

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 04:58:59, Search time 4328 Seconds

(without alignments)
11036.050 Million cell updates/sec

Title: US-10-037-417-45

Sequence: 1 999ccctcgtcctcgtggtccat.....999gttcgtatggggtcc 1102

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : GenBank1:
1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pla:*
35: em_hg_rod:*
36: em_hg_mam:*
37: em_hg_vrt:*
38: em_sy:*
39: em_hg_hum:*
40: em_hg_mus:*
41: em_hg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1102	100.0	1102	6	AX512287
2	1102	100.0	1102	6	AX512289
3	815.6	74.0	2457	6	AX360096
4	815.6	74.0	2681	6	AX480935
5	806.6	73.2	2810	6	BD127529
6	806.6	73.2	2810	9	AK075142
7	535.4	48.6	537	6	AX342934
8	472.8	42.9	670	6	BD125219
9	472.8	42.9	670	6	BD126485
10	442.8	40.2	456	6	AX342936
11	280.2	25.4	127769	9	AC009088
12	280.2	25.4	195476	2	AC135044
13	232.4	21.1	1796	6	AX098193
14	232.4	21.1	1800	9	BC001462
15	232.4	21.1	1835	6	AX098215
16	226.4	20.5	944	6	AX375744
17	225.6	20.5	1020	6	AX823252
18	225.6	20.5	1718	9	BC062334
19	225.6	20.5	1755	9	BC036846
20	224	20.3	1606	6	AX676264
21	224	20.3	1613	6	AR256990
22	224	20.3	1613	9	AF536382
23	221.2	20.1	1834	6	AX335777
24	221.2	20.1	1834	6	AX336076
25	221.2	20.1	1834	6	AX336540
26	221.2	20.1	1834	6	AX409567
27	221.2	20.1	1834	6	AX474697
28	221.2	20.1	1834	9	HUMPROS
29	220.2	20.0	71177	2	AC101465
30	220.2	20.0	185788	10	AC124461
31	220.2	20.0	219763	10	AC093175
32	217.8	19.8	232119	2	AC106629
33	217.8	19.8	244161	2	AC117170
34	217	19.7	64328	2	AC124816
35	216.8	19.7	1130	6	AR256996
36	205.2	18.6	786	6	AX375740
37	200	18.1	1727	10	AX335911
38	200	18.1	1743	10	BC003851
39	200	18.1	1856	10	AB038244
40	200	18.1	1870	10	AF378085
41	195.2	17.7	1797	10	AF188613
42	187.8	17.0	1360	10	AY262280
43	185.6	16.8	882	6	AX675583
44	183.2	16.6	1142	6	AR219285
45	183.2	16.6	1169	6	AR219284

ALIGNMENTS

RESULT 1
AX512287
LOCUS
DEFINITION
AX512287
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Kekuda, R., Alsobrook, J. P., Tcherny, V. T., Liu, X., Sytek, K. A.,
Patturajan, M., Grosse, W. M., Lepley, D. M., Burgess, C. B., Verne, C. A.,
Li, L., Gorman, L., Edinger, S., Sciore, P., Ellerman, K., Malyankar, V.,

Sequence 43 from Patent WO02053742. DNA linear PAT 27-SEP-2002
AX512287.1 GI:23392631

Rothenberg, M., Stone, D., Boldog, F., Shenoy, S., and Anderson, D.
 Proteins and nucleic acids encoding same
 Patent: WO 02053742-A 43 11-JUL-2002;
 Curagen Corporation (US)
 FEATURES
 source
 1. .1102
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1102; DB 6; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 2.6e-164;
 Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTCTCTCTGGGCCATGGCCAGAGAGGGGCTCTGGGGCCCTGGGGCAGCTGGGGGCT 60
 DB 1 GGGCCCTCTCTCTGGGCCATGGCCAGAGAGGGGCTCTGGGGCCCTGGGGCAGCTGGGGGCT 60

QY 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTTGGTGCCTCCGGACCCCGCTAGGGGC 120
 DB 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTTGGTGCCTCCGGACCCCGCTAGGGGC 120

QY 121 CCGCGTACTGCGGGCCCTGAGCCCTCGGCCGATCGTGGGGGCTCAAACCGCGAG 180
 DB 121 CCGCGTACTGCGGGCCCTGAGCCCTCGGCCGATCGTGGGGGCTCAAACCGCGAG 180

QY 181 CCGGGCACTGGCTTGGCAAGTGGAGCTGCACCATGGAGTGGCCACATCTGCGGGGC 240
 DB 181 CCGGGCACTGGCTTGGCAAGTGGAGCTGCACCATGGAGTGGCCACATCTGCGGGGC 240

QY 241 TCCCTCATGCCCCCTCTGGGCTCTCTCCGTCTCTCCGTCTCACTTTTATGACGAATGGGACG 300
 DB 241 TCCCTCATGCCCCCTCTGGGCTCTCTCCGTCTCTCCGTCTCACTTTTATGACGAATGGGACG 300

QY 301 TTGGAGCCCGCGGCGAGTGGTGGTCTGCTGGGCTGCACTCCGAGACGGGCGCCCTG 360
 DB 301 TTGGAGCCCGCGGCGAGTGGTGGTCTGCTGGGCTGCACTCCGAGACGGGCGCCCTG 360

QY 361 GACGGCGGCGACACCGCGGCGAGTGGCTGCTGGGCTGCGGGCCAACTACAGCCAACTG 420
 DB 361 GACGGCGGCGACACCGCGGCGAGTGGCTGCTGGGCTGCGGGCCAACTACAGCCAACTG 420

QY 421 GAGCTGGGCGCGACCTGGCCCTGCTGGGCTGCGGCTCAACCGCGAGCTGGGCGCCGCG 480
 DB 421 GAGCTGGGCGCGACCTGGCCCTGCTGGGCTGCGGCTCAACCGCGAGCTGGGCGCCGCG 480

QY 481 GTGTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
 DB 481 GTGTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540

QY 541 GCCACCGGCTGGGGAGAGCTCCAGAGGAGATCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 600
 DB 541 GCCACCGGCTGGGGAGAGCTCCAGAGGAGATCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 600

QY 601 GAGTGGAGCTAAGCTGCTGGCGAGGCACTCTCAATGCTCTAAGCAGCCCGCTG 660
 DB 601 GAGTGGAGCTAAGCTGCTGGCGAGGCACTCTCAATGCTCTAAGCAGCCCGCTG 660

QY 661 CCGTTCACCTCACTCTCCAGATATGCGAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 720
 DB 661 CCGTTCACCTCACTCTCCAGATATGCGAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 720

QY 721 CGCAGGACACCTGCGAGGGTGAATCTGGGGGGCCCTGGTCTGTGAGGAGGCGGGCGCG 780
 DB 721 CGCAGGACACCTGCGAGGGTGAATCTGGGGGGCCCTGGTCTGTGAGGAGGCGGGCGCG 780

QY 781 TGGTTCAGAGGAGGAATCAACAGCTTTGGGTTGGTGTGAGACGAGAAAACCGCCCTGGA 840
 DB 781 TGGTTCAGAGGAGGAATCAACAGCTTTGGGTTGGTGTGAGACGAGAAAACCGCCCTGGA 840

QY 841 GTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 DB 841 GTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900

841 GTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 QY 901 CTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
 DB 901 CTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960

QY 961 GCATTCCTGGATCTGCGCAGAACTCTTTTGGGCCCTGCTCCATATATCAGTAGAGTC 1020
 DB 961 GCATTCCTGGATCTGCGCAGAACTCTTTTGGGCCCTGCTCCATATATCAGTAGAGTC 1020

QY 1021 TCAACTGGGACCAAAAGCCTTGTCTCCCTCCCTGGCTCTCTCCACACTCTCTCTCTGGGCTC 1080
 DB 1021 TCAACTGGGACCAAAAGCCTTGTCTCCCTCCCTGGCTCTCTCCACACTCTCTCTCTGGGCTC 1080

QY 1081 TGGGGGCTCTGATGGGGGCTCC 1102
 DB 1081 TGGGGGCTCTGATGGGGGCTCC 1102

RESULT 2
 AX512289
 LOCUS AX512289 1102 bp DNA linear PAT 27-SEP-2002
 DEFINITION Sequence 45 from Patent WO02053742.
 ACCESSION AX512289
 VERSION AX512289.1 GI:23392632
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1
 AUTHORS
 Kekula, R., Alsobrook, J.P., Tchernev, V.T., Liu, X., Spytek, K.A.,
 Patturajan, M., Grosse, W.M., Lepley, D.M., Burgess, C.E., Vernet, C.A.,
 Li, J., Gorman, L., Edinger, S., Sciore, P., Ellerman, K., Malyankar, U.,
 Rothenberg, M., Stone, D., Boldog, F., Shenoy, S. and Anderson, D.
 Proteins and nucleic acids encoding same
 Patent: WO 02053742-A 45 11-JUL-2002;
 Curagen Corporation (US)

FEATURES
 Location/Qualifiers
 1. .1102
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1102; DB 6; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 2.6e-164;
 Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTCTCTCTGGGCCATGGCCAGAGAGGGGCTCTGGGGCCCTGGGGCAGCTGGGGGCT 60
 DB 1 GGGCCCTCTCTCTGGGCCATGGCCAGAGAGGGGCTCTGGGGCCCTGGGGCAGCTGGGGGCT 60

QY 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTTGGTGCCTCCGGACCCCGCTAGGGGC 120
 DB 61 GTGGCCAAATCTGACTCATATCACTTTACGGGTTGGTGCCTCCGGACCCCGCTAGGGGC 120

QY 121 CCGCGTACTGCGGGCCCTGAGCCCTCGGCCGATCGTGGGGGCTCAAACCGCGAG 180
 DB 121 CCGCGTACTGCGGGCCCTGAGCCCTCGGCCGATCGTGGGGGCTCAAACCGCGAG 180

QY 181 CCGGGCACTGGCTTGGCAAGTGGAGCTGCACCATGGAGTGGCCACATCTGCGGGGC 240
 DB 181 CCGGGCACTGGCTTGGCAAGTGGAGCTGCACCATGGAGTGGCCACATCTGCGGGGC 240

QY 241 TCCCTCATGCCCCCTCTGGGCTCTCTCCGTCTCTCCGTCTCACTTTTATGACGAATGGGACG 300
 DB 241 TCCCTCATGCCCCCTCTGGGCTCTCTCCGTCTCTCCGTCTCACTTTTATGACGAATGGGACG 300

QY 301 TTGGAGCCCGCGGCGAGTGGTGGTCTGCTGGGCTGCACTCCGAGACGGGCGCCCTG 360
 DB 301 TTGGAGCCCGCGGCGAGTGGTGGTCTGCTGGGCTGCACTCCGAGACGGGCGCCCTG 360

QY 361 GACGGCGGCGACACCGCGGCGAGTGGCTGCTGGGCTGCGGGCCAACTACAGCCAACTG 420
 DB 361 GACGGCGGCGACACCGCGGCGAGTGGCTGCTGGGCTGCGGGCCAACTACAGCCAACTG 420

QY 421 GAGCTGGGCGCGACCTGGCCCTGCTGGGCTGCGGCTCAACCGCGAGCTGGGCGCCGCG 480
 DB 421 GAGCTGGGCGCGACCTGGCCCTGCTGGGCTGCGGCTCAACCGCGAGCTGGGCGCCGCG 480

QY 481 GTGTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
 DB 481 GTGTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540

QY 541 GCCACCGGCTGGGGAGAGCTCCAGAGGAGATCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 600
 DB 541 GCCACCGGCTGGGGAGAGCTCCAGAGGAGATCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 600

QY 601 GAGTGGAGCTAAGCTGCTGGCGAGGCACTCTCAATGCTCTAAGCAGCCCGCTG 660
 DB 601 GAGTGGAGCTAAGCTGCTGGCGAGGCACTCTCAATGCTCTAAGCAGCCCGCTG 660

QY 661 CCGTTCACCTCACTCTCCAGATATGCGAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 720
 DB 661 CCGTTCACCTCACTCTCCAGATATGCGAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 720

QY 721 CGCAGGACACCTGCGAGGGTGAATCTGGGGGGCCCTGGTCTGTGAGGAGGCGGGCGCG 780
 DB 721 CGCAGGACACCTGCGAGGGTGAATCTGGGGGGCCCTGGTCTGTGAGGAGGCGGGCGCG 780

QY 781 TGGTTCAGAGGAGGAATCAACAGCTTTGGGTTGGTGTGAGACGAGAAAACCGCCCTGGA 840
 DB 781 TGGTTCAGAGGAGGAATCAACAGCTTTGGGTTGGTGTGAGACGAGAAAACCGCCCTGGA 840

QY 841 GTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 DB 841 GTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900

QY 361 GAGCGCGGCACACCCGCGCAGTGGCGGCCATCGTGGTCCGCGCCAACTACAGCCAAAGTG 420
Db 361 GAGCGCGGCACACCCGCGCAGTGGCGGCCATCGTGGTCCGCGCCAACTACAGCCAAAGTG 420
QY 421 GAGCTGGGCGCGCACTGGCCCTGCTGGGCTGGGCTCTACCCGCGCAGCTGGGCGCCCGCC 480
Db 421 GAGCTGGGCGCGCACTGGCCCTGCTGGGCTGGGCTCTACCCGCGCAGCTGGGCGCCCGCC 480
QY 481 GTGTGGCTGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db 481 GTGTGGCTGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
QY 541 GGCACCGGTGGGAGACGTCGAGAGGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 541 GGCACCGGTGGGAGACGTCGAGAGGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
QY 601 GAAGTGGAGTAAAGCTGCTGGGCGAGGACCTGTCAATGCTCTACAGCCAGCCCGGT 560
Db 601 GAAGTGGAGTAAAGCTGCTGGGCGAGGACCTGTCAATGCTCTACAGCCAGCCCGGT 560
QY 661 CCCTTCAACCTCACTCTCCAGATATTGCGAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 661 CCCTTCAACCTCACTCTCCAGATATTGCGAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 720
QY 721 CGAGGAGACACCTGCGAGGCTGCTGCGGCGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 721 CGAGGAGACACCTGCGAGGCTGCTGCGGCGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 781 TGGTTCCAGCGAGGATACCGAGCTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTT 840
Db 781 TGGTTCCAGCGAGGATACCGAGCTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTT 840
QY 841 GTTTTCACTGCTGTGCTACCTATGAGGATGATACGGGAGCAGGTGATGGTTTCAGAG 900
Db 841 GTTTTCACTGCTGTGCTACCTATGAGGATGATACGGGAGCAGGTGATGGTTTCAGAG 900
QY 901 CTGGGCGTGCCTTTCCACCCAGCCCGCAGAGACCCAGTCAAGTGTGTTTATCAAAAG 960
Db 901 CTGGGCGTGCCTTTCCACCCAGCCCGCAGAGACCCAGTCAAGTGTGTTTATCAAAAG 960
QY 961 GCATTCTCGATTCTGCCAGATCCCTTTGAGGCGCTTCTCCCATATATACGTAGAGTTC 1020
Db 961 GCATTCTCGATTCTGCCAGATCCCTTTGAGGCGCTTCTCCCATATATACGTAGAGTTC 1020
QY 1021 TCACTGGGACCAAAAGCCTTGTCTCCCTGCTGCTCTCTCCACACTCTCTCTCTGCGCCTC 1080
Db 1021 TCACTGGGACCAAAAGCCTTGTCTCCCTGCTGCTCTCTCTCCACACTCTCTCTCTGCGCCTC 1080
QY 1081 TGGGGTTCTGATGGGCGCTCC 1102
Db 1081 TGGGGTTCTGATGGGCGCTCC 1102

RESULT 3
AX360096
LOCUS
DEFINITION
Sequence 52 from Patent WO0200860.
AX360096
ACCESSION
VERSION
AX360096.1 GI:18675722
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Plozman, G., Whyte, D., Sudarsanam, S., Manning, G., Caenepeel, S. and
Charyczak, G.
TITLE
Novel proteases
JOURNAL
Patent: WO 0200860-A 52 03-JAN-2002;
Sugen, Inc. (US)
FEATURES
Location/Qualifiers
1..2457
/organism="Homo sapiens"

/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 74.0%; Score 815.6; DB 6; Length 2457;
Best Local Similarity 99.5%; Pred. No. 3.1e-119;
Matches 818; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 124 CGTACTGCGGCGCCCTGAGCCCTCGGCCCATCGTGGGGGGCTCAACCGCGAGCCG 183
Db 106 CTGGAGTCTGGGCGCCCTGAGCCCTCGGCCCATCGTGGGGGGCTCAACCGCGAGCCG 165
QY 184 GGCACCTGGCCTTGGCAAGTGAAGCTGACCATGGAGGTGGCCACATCTGCGGGGCTCC 243
Db 166 GGCACCTGGCCTTGGCAAGTGAAGCTGACCATGGAGGTGGCCACATCTGCGGGGCTCC 225
QY 244 CTCACTGGCCCTCTGGGTCTCTCGCTGCTCACTGTTTCAATGAGATGGAGCTTG 303
Db 226 CTCACTGGCCCTCTGGGTCTCTCGCTGCTCACTGTTTCAATGAGATGGAGCTTG 285
QY 304 GAGCCCGCGCGAGTGGTTCGCTACTGCTGGCGTGCATCCACGAGGAGGGGCCCTGGAC 363
Db 286 GAGCCCGCGCGAGTGGTTCGCTACTGCTGGCGTGCATCCACGAGGAGGGGCCCTGGAC 345
QY 364 GCGCGGCACACCCGCGCAGTGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 423
Db 346 GCGCGGCACACCCGCGCAGTGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 405
QY 424 CTGGGCGCGGACCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
Db 406 CTGGGCGCGGACCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 465
QY 484 TGGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543
Db 466 TGGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
QY 544 ACCGCTGCGGAGAGCTCCAGGAGCAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 603
Db 526 ACCGCTGCGGAGAGCTCCAGGAGCAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 585
QY 604 GTGGAGCTAAGGCTGCTGGGCGAGGCGCCTGCTCAATGCTCTCTACAGCCAGCCCGTCC 663
Db 586 GTGGAGCTAAGGCTGCTGGGCGAGGCGCCTGCTCAATGCTCTCTACAGCCAGCCCGTCC 645
QY 664 TTCAACCTCACCTCCAGATATTGCCAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 723
Db 646 TTCAACCTCACCTCCAGATATTGCCAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 705
QY 724 AGGGACACCTGCCAGGGTGAATCTGGGGGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 783
Db 706 AGGGACACCTGCCAGGGTGAATCTGGGGGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765
QY 784 TTCCAGGCGAGGATCACAGCTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTT 843
Db 766 TTCCAGGCGAGGATCACAGCTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTT 825
QY 844 TTCACTGCTGTGGTACCTATGAGGATGATGATGAGGATGATGATGAGGATGATGATGAGGATGATGATG 903
Db 826 TTCACTGCTGTGGTACCTATGAGGATGATGATGAGGATGATGATGAGGATGATGATGAGGATGATGATG 885
QY 904 GGGCCTGCTTTCCACCCAGCCCGCAGAGCCCGCAGATCAGAT 945
Db 886 GGGCCTGCTTTCCACCCAGCCCGCAGAGCCCGCAGATCAGAT 927
RESULT 4
AX480935
LOCUS
DEFINITION
Sequence 31 from Patent WO0246383.
AX480935
ACCESSION
VERSION
AX480935.1 GI:22217574
KEYWORDS
SOURCE
Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Yue H., Azimzai, Y., Kallick D.A., Baughn M.R., Griffin, J.A.,
Swarnakar, A., Lal P.G., Walla, N.K., Hafalia, A.J., Gandhi, A.R.,
Au-Young, J., Elliott, V.S., Ramkumar, J., Thangavelu, K., Lu, Y.,
Warren, B.A., Lu, D.A., Lee, E.A., Tribouley, C.M., Arvizu, C.,
Delegeane, A.M., Yao, M.G., Khan, F.A. and Sanjanwala, M.M.
TITLE Protein modification and maintenance molecules
JOURNAL Patent: WO 0246383-A 31.13-JUN-2002;
Incyte Genomics, Inc. (US)
FEATURES
source Location/Qualifiers
1..2881
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 2751509CB1"
ORIGIN
Query Match 74.0%; Score 815.6; DB 6; Length 2681;
Best Local Similarity 99.5%; Pred. No. 3e-119;
Matches 818; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 124 CGGTACTGGGGCGCCCTGAGCCCTGGCCCGCATCGTGGGGGCTCAAAACGCGAGCGG 183
Db 106 CTGGACTGGGGCGCCCTGAGCCCTGGCCCGCATCGTGGGGGCTCAAAACGCGAGCGG 165
QY 184 GGCACCTGGCCCTTGGCAAGTGGAGCTGCACCTGAGAGTGGCCACATCTGCGGGGCTCC 243
Db 166 GGCACCTGGCCCTTGGCAAGTGGAGCTGCACCTGAGAGTGGCCACATCTGCGGGGCTCC 225
QY 244 CTCATCGCCCTCTCTGGGTCTCTCCGGTGTCTCATCTGTTTCATGACGAATGGAGCGTTG 303
Db 226 CTCATCGCCCTCTCTGGGTCTCTCCGGTGTCTCATCTGTTTCATGACGAATGGAGCGTTG 285
QY 304 GAGCCCGCGCGAGTGGTCTGCTGCTGGCGTGCACCTCCAGAGCGGGCCCTTGAC 363
Db 286 GAGCCCGCGCGAGTGGTCTGCTGCTGGCGTGCACCTCCAGAGCGGGCCCTTGAC 345
QY 364 GCGCGGCACACCCGCGCAGTGGCGCATCTGCTGGTGGCGGCACTACAGCAAGTGGAG 423
Db 346 GCGCGGCACACCCGCGCAGTGGCGCATCTGCTGGTGGCGGCACTACAGCAAGTGGAG 405
QY 424 CTGGGGCGCGACTGGCCCTCTGCGCTGGCTGACCTACCGCGAGCCCTGGCCCGCTG 483
Db 406 CTGGGGCGCGACTGGCCCTCTGCGCTGGCTGACCTACCGCGAGCCCTGGCCCGCTG 465
QY 484 TGGCCTGTGCTGCTGCGCGCCCTCACACCGCTTCTGTCAGCGCACCGCTGCGGCC 543
Db 466 TGGCCTGTGCTGCTGCGCGCCCTCACACCGCTTCTGTCAGCGCACCGCTGCGGCC 525
QY 544 ACCGCTGGGGAGAGCGTCAGAGGAGAGATCTCTGCTCTCCCTGCGTGTACAGGAA 603
Db 526 ACCGCTGGGGAGAGCGTCAGAGGAGAGATCTCTGCTCTCCCTGCGTGTACAGGAA 585
QY 604 GTGGAGCTAAGCTGCTGGCGAGGCCACCTGCTCAATGCTCTTACAGCCAGCCCGTCC 663
Db 586 GTGGAGCTAAGCTGCTGGCGAGGCCACCTGCTCAATGCTCTTACAGCCAGCCCGTCC 645
QY 664 TTCAACCTCACTCTCCAGATATGCGAGGATGCTGTGTGCTGCTGCTGCTGCTGCTG 723
Db 646 TTCAACCTCACTCTCCAGATATGCGAGGATGCTGTGTGCTGCTGCTGCTGCTGCTG 705
QY 724 AGGGACCTGCGCAGGTGACTCTGGGGGGCCCTTGGTGTGTGAGAAAGGGCGCGCTGG 783
Db 706 AGGGACCTGCGCAGGTGACTCTGGGGGGCCCTTGGTGTGTGAGAAAGGGCGCGCTGG 765
QY 784 TTCCAGGAGGAGTACACAGCTTTGGGTTTGGCTGTGACGAGAGAAACCGCCCTGAGTT 843
Db 766 TTCCAGGAGGAGTACACAGCTTTGGGTTTGGCTGTGACGAGAGAAACCGCCCTGAGTT 825
844 TTCACTGTGTGGCTACCTATGAGGCATGGATACGGGAGAGGTGATGGGTTTCAGAGCCT 903

Db 826 TTCCTGTGTGGCTACCTATGAGCATGATCGGGAGCATGATGGTTTCAGAGCCT 885
QY 904 GGGCCTGCTTTCCACCAGCCCGCCAGAGACCCAGTCAGAT 945
Db 886 GGGCCTGCTTTCCACCAGCCCGCCAGAGACCCAGTCAGAT 927
RESULT 5
BD127529 2810 bp DNA linear PAT 18-SEP-2002
LOCUS Primer for synthesizing full-length cDNA and use thereof.
DEFINITION
ACCESSION BD127529
VERSION BD127529.1 GI:23224474
KEYWORDS JP 2002017375-A/2960.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2810)
Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
Koga, H.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
PN JP 2002017375-A/2960
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUO OTSUKI, HISASHI KOGA
PC
PC N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
PC C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
FT CDS Location/Qualifiers
1..2810
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 73.2%; Score 806.6; DB 6; Length 2810;
Best Local Similarity 92.0%; Pred. No. 7.9e-118;
Matches 862; Conservative 0; Mismatches 74; Indels 1; Gaps 1;
QY 10 TCCTGGGCGATGGCCAGAGGGGCTCTGGGGCTGGGAGCTGGGGCTGTGGCCAT 69
Db 23 TCTAGCCATGGCCCGGACCTGCTCTCCCTTGTGATGCTGTGATCAGTCCATC 82
QY 70 TCTGACTCATACTCACTTTACGGGTGTGTGCTGCGGACCGCTAGGGGCGCCCGGTAC 129
Db 83 CCAGGAGCTTCCAGGACTCAGCTCTCAGTCTTACCAGGAGAACCTGAAGATCTGGAC 142
QY 130 TGGGGGCGCTGAGCCCTGGCCCGCATCTGT-TGGGGGCTCAAAACGCGAGCGGGGAC 188
Db 143 TGGGGGCGCTGAGCCCTGGCCCGCATCTGTGGGGGGGCTCAAAACGCGAGCGGGGAC 202
QY 189 CTGGCTTGGCAAGTGAAGCTGACCTGAGAGTGGCCACATCTGCGGGGGCTCCCTCAT 248
Db 203 CTGGCTTGGCAAGTGAAGCTGACCTGAGAGTGGCCACATCTGCGGGGGCTCCCTCAT 262
QY 249 CGCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTTCATGACGAATGGAGCTTGGAGCC 308
Db 263 CGCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTTCATGACGAATGGAGCTTGGAGCC 322

```
QY 309 CGCGCCGAGTGGTGGTACTGCTGGGCGTGCATCCACGAGCGGGCCCTCTGACGGCGC 368
Db 323 CGCGCCGAGTGGTGGTACTGCTGGGCGTGCATCCACGAGCGGGCCCTCTGACGGCGC 382
QY 369 GCACACCCGCGAGTGGCGCCCATGCTGGTGGCGGCACTACAGCAAGTGGAGCTGGG 428
Db 383 GCACACCCGCGAGTGGCGCCCATGCTGGTGGCGGCACTACAGCAAGTGGAGCTGGG 442
QY 429 CGCGACCTGGCCCTGCTGGGCGTGGCTCAACCGCGAGCCTGGGCGCCCGCGTGGGCG 488
Db 443 CGCGACCTGGCCCTGCTGGGCGTGGCTCAACCGCGAGCCTGGGCGCCCGCGTGGGCG 502
QY 489 TGCTGCTGCTGGCCCGGCTCTACACCGCTTGTGTCAGCGGACCGCTGCTGGGCGACCGG 548
Db 503 TGCTGCTGCTGGCCCGGCTCTACACCGCTTGTGTCAGCGGACCGCTGCTGGGCGACCGG 562
QY 549 CTGGGAGAGCTCCAGGAGCGAGATCTCTGCTGCTCTCCCTGGGTGCTACAGGAGTGA 608
Db 563 CTGGGAGAGCTCCAGGAGCGAGATCTCTGCTGCTCTCCCTGGGTGCTACAGGAGTGA 622
QY 609 GCTAAGGCTGCTGGGCGAGGCCACCTGTCAATGTCTCTACAGCGACCGCGTCCCTTCAA 668
Db 623 GCTAAGGCTGCTGGGCGAGGCCACCTGTCAATGTCTCTACAGCGACCGCGTCCCTTCAA 682
QY 669 CTTCACTCTCCAGATATTCGCAAGGATGCTGTGCTGTCAGCGGACCGCTGCTGGGCGACCGG 728
Db 683 CTTCACTCTCCAGATATTCGCAAGGATGCTGTGCTGTCAGCGGACCGCTGCTGGGCGACCGG 742
QY 729 CACTGTCAGGCTGACTCTGGGCGGCGCTGTCTGTGAGGAAGCGCGCTGCTGCTCA 788
Db 743 CACTGTCAGGCGACTCTGGGCGGCGCTGTGTGAGGAAGCGCGCTGCTGCTCA 802
QY 789 GGCAGGATCACCAGCTTTGGTGTGGTGTGGACGAGAAACCGCCCTGGAGTTTTCAC 848
Db 803 GGCAGGATCACCAGCTTTGGTGTGGTGTGGACGAGAAACCGCCCTGGAGTTTTCAC 862
QY 849 TGCTGTGGCTACTATGAGGATGATAGGAGGAGAGTGTGAGGTTACAGCCTGGGCG 908
Db 863 TGCTGTGGCTACTATGAGGATGATAGGAGGAGAGTGTGAGGTTACAGCCTGGGCG 922
QY 909 TGCCCTTCCACCCAGCCCGCAGAGACCGACGTCAGAT 945
Db 923 TGCCCTTCCACCCAGCCCGCAGAGACCGACGTCAGAT 959

RESULT 6
AK075142
LOCUS
DEFINITION
Homo sapiens cDNA FLJ90661 fis, clone PLACE1005003, weakly similar
to PROSTASIN PRECURSOR (EC 3.4.21.-).
ACCESSION
AK075142
VERSION
AK075142.1 GI:22761040
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Isogai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T.,
Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S.,
Kawai-Hio, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y.,
Kojima, S., Nagahari, K., Masuho, Y., Ono, T., Okano, K., Yoshikawa, Y.,
Aotsuma, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and
Ninomiya, K.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2810)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
```

Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).

FEATURES

source

1..2810
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1005003"
/tissue_type="placenta"
/clone_lib="PLACE1"
/note="cloning vector: pME18SFL3"
300..2600
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC11431.1"
/db_xref="GI:22761041"

CDS

/translation="MTNCTLEPAEWSVLLGVHSDGPDLDGAHTFAVAIVPANYSQ
VELGADLALLSLASPASLGPAWVPCLPASHRFVHGTAACNATGWDVQEDAPLELPW
VLQVELKLGATCCCLYSQPGPNLTLLQILPGMLCAGYEGRRDTCQSGSGPLVC
REGDFOAGITTSFGCGRRNRPGVFATAVYEAWIREQVMSSEPFAPTPQKQTCF
SDQPRENCTIALPECGKAPRPAQVAVMVRGSRPCHGALVSESWLAPASCF
LDPNSSDPPRDLDAWRVLLPSRAERVARLVHENAASWDNASDLALLQLTPNLS
AASRPVCLPHEHYFLPGSRCLARWGRGEPALGEGALLAEALLGGWCHCLYGRGA
AVPLQTHGPVSHVTRGAYLEDQLADWGPDBGETETCTCPPTTHSHGACGURLEAPV
FVLMFWLAWHVAGDVCTGILLAPWVLAATHCVLRFGSTTPVYVYLRAGASLL
PQGHQSLRVSIRLPHQLRPLALLESVPESFALPICLHPAGIPPGASCWV
LGWKEPDVRPVAASVILTORICDLVQGLTPGTLCLVAEGDENCMTSAPLL
COMTEGSLVGMVAVQSGRRELFALGPPEANISQTVGEANLLPPSPSPHWPVGGSNLC
PPELAKAGSPHAYFLLLLTLLIQS"

ORIGIN

Query Match 73.2%; Score 806.6; DB 9; Length 2810;
Best Local Similarity 92.0%; Pred. No. 7.9e-118;
Matches 862; Conservative 0; Mismatches 74; Indels 1; Gaps 1;
QY 10 TCCTGGGCGCATGGCCAGAAAGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCAAAT 69
Db 23 TCTAGCGCATGGCCCGGCACCTGCTCTCCCTCTTGTGATGCTTGTCCATCAGTCCCATC 82
QY 70 TCTGACTCATCTCACTTACGGGTGGTGGCGTCCGACCCGCTAGGGGCGCCCGGTAC 129
Db 83 CCAGGAGCTTCCAGGACTCAGCTCAGTCTCAGTCTACCCAGGAGAACCTGAAGATCTGAC 142
QY 130 TCGGGGCGCCTGAGCCCTCGGCGCGCATGCT-GGGGGGCTCAAAACGCGAGCGGGGCAC 188
Db 143 TCGGGGCGCCTGAGCCCTCGGCGCGCATGCTGGGGGGGCTCAAAACGCGAGCGGGGCAC 202
QY 189 CTGCGCTTGGCAAGTGAAGCTGACCAATGAGGTGGCCACATCTCGGGGGCTCCCTCAT 248
Db 203 CTGGCTTGGCAAGTGAAGCTGACCAATGAGGTGGCCACATCTCGGGGGGCTCCCTCAT 262
QY 249 CGCCCTCTCTCGGTCTCTCCGCTGCTCACTGTTCATGACGAATGGCACTGTGGAGCC 308
Db 263 CGCCCTCTCTCGGTCTCTCTCCGCTGCTCACTGTTCATGACGAATGGCACTGTGGAGCC 322
QY 309 CGCGGCGAGTGTGGTGTGCTGCTGGGGGTGCACTCCCGAGAGCGGGCCCTTGAGCGGGC 368
Db 323 CGCGGCGAGTGTGGTGTGCTGCTGGGGGTGCACTCCCGAGAGCGGGCCCTTGAGCGGGC 382
QY 369 GCACACCGCGCAGTGGCGCCCATCTGCTGGTGGCGGCCCACTACAGCCCAAGTGGAGCTGGG 428
Db 383 GCACACCGCGCAGTGGCGCCCATCTGCTGGTGGCGGCCCACTACAGCCCAAGTGGAGCTGGG 442
QY 429 CGCGGACCTGGCCCTGCTGGCGCTTGGGCTCACCGCGCGAGCTGGGGCCCGCGGTGTGGCC 488
Db 443 CGCGGACCTGGCCCTGCTGGCGCTTGGGCTCACCGCGCGAGCTGGGGCCCGCGGTGTGGCC 502
QY 489 TGTCTGCTGCCCCCGCGCTCACACCGCTTGTGTGCAAGGACCGCCCTGTGGGCCACCGG 548


```

70 TCTGACTCATCTACCTTTACGGTGTGTGCGTCCGGACCGCTAGGGGCCCCCGCTAC 129
Db CCAGGAGCCTCCAGGACTCAGCTCTCAGTCTACTCCAGGAAGACCTGAAGATCTGGAC 142
QY 130 TCGGGGCGCCTGAGCCCTCGGCCCGCATCGT- GGGGGGCTCAAAACGGCGAGCCGGGCAC 188
Db 143 TCGGGGCGCCTGAGCCCTCGGCCCGCATCGTGGGGGGGCTCAAAACGGCGAGCCGGGCAC 202
QY 189 CTGGCTTTGGCAAGTGAGCTGCACCATGAGAGTGGCCACATCTCGGGGGCTCCCTCAT 248
Db 203 CTGGCTTTGGCAAGTGAGCTGCACCATGAGAGTGGCCACATCTCGGGGGCTCCCTCAT 262
QY 249 CGCCCCCTCTCGGGCTCTCGCGCTCTCAGTGTTCATGACGAATGGAGCCTTGGAGCC 308
Db 263 CGCCCCCTCTCGGGCTCTCGCGCTCTCAGTGTTCATGACGAATGGAGCCTTGGAGCC 322
QY 309 CGCGCCGAGTGGTCTGCTACTCTGGGGTGCATCTCCAGGACGGGCCCTTGGAGCGGCG 368
Db 323 CGCGCCGAGTGGTCTGCTACTCTGGGGTGCATCTCCAGGACGGGCCCTTGGAGCGGCGC 382
QY 369 GCACACCGCGCAGTGGCGGCCATCGTGTGCGCGCCAACTACAGCAAGTGGAGCTGG 428
Db 383 GCACACCGCGCAGTGGCGGCCATCGTGTGCGCGCCAACTACAGCAAGTGGAGCTGG 442
QY 429 CGCGACCTGGCCCTCTGCGCTCTCAGTGTTCATGACGAATGGAGCCTTGGAGCC 488
Db 443 CGCGACCTGGCCCTCTGCGCTCTCAGTGTTCATGACGAATGGAGCCTTGGAGCC 502
QY 489 TGTCTGCTGCCCCCGCGCTCAGCAGCGCTTGTGTCAGCGCACCGCTGCTGGGCCACCGG 548
Db 503 TGTCTGCTGCCCCCGCGCTCAGCAGCGCTTGTGTCAGCGCACCGCTGCTGGGCCACCGG 562
QY 549 CTGGGAGACGCTCCA- GGAGGACAGATCTCTGCTCTCCCTGGGTGCTAC- AGGAAGTG 606
Db 563 CTGGGAGACGCTCCA- GGAGGAGACGATCTCTGCTCTCCCTGGGTGCTAC- AGGAAGTG 622
QY 607 GAGCTAAGGCTCTGGGCG- AGGCCACCTGTCAATGTCTTACAGCCA 653
Db 623 GAGCTAAGGCTCTGGGCGAAGGCAACTGGTCAATGTCTTACANNCA 670

RESULT 9
BD126485 670 bp DNA linear PAT 18-SEP-2002
LOCUS
DEFINITION
ACCESSION
BD126485
VERSION
BD126485.1 GI:23221430
KEYWORDS
JP 2002017375-A/1916.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 1916 22-JAN-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/1916
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOgai,KOJI HAYASHI,SHIZUKO
PI ISHII,
PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI,HISASHI KOGA
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10, C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
```

```

FEATURES
source
FT source 1..670
FT /organism='Homo sapiens (human)'
Location/Qualifiers
1..670
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Query Match 42.9%; Score 472.8; DB 6; Length 670;
Best Local Similarity 86.9%; Pred. No. 4.9e-65;
Matches 563; Conservative 0; Mismatches 81; Indels 4; Gaps 4;
QY 10 TCCTGGGCGCATGGCCAGAGGGGCTCTGGGGGCTGGGAGCTGGGGGCTGTGGCCAT 69
Db 23 TCTAGCGCATGGCCCGGCACCTGCTCTCCCTTGTGATGCTTGTCAATCAGTCCCCATC 82
QY 70 TCTGACTCATCTCATTACGGGTTGGTGGCGTGGGACCGCTAGGGGGCCCCCGCTAC 129
Db 83 CCAGGAGCCTTCCAGGACTCAGTCTCAGTCTCTACCCAGGAAGAACCTGAAGATCTGGAC 142
QY 130 TCGGGGCGCCTGAGCCCTCGGCCCGCATCGT- GGGGGGCTCAAAACGGCGAGCCGGGCAC 188
Db 143 TCGGGGCGCCTGAGCCCTCGGCCCGCATCGTGGGGGGGCTCAAAACGGCGAGCCGGGCAC 202
QY 189 CTGGCTTTGGCAAGTGAGCTGCACCATGAGAGTGGCCACATCTCGGGGGCTCCCTCAT 248
Db 203 CTGGCTTTGGCAAGTGAGCTGCACCATGAGAGTGGCCACATCTCGGGGGCTCCCTCAT 262
QY 249 CGCCCCCTCTCGGGTCTCTCGCTGCTCAGTGTTCATGACGAATGGAGCCTTGGAGCC 308
Db 263 CGCCCCCTCTCGGGTCTCTCGCTGCTCAGTGTTCATGACGAATGGAGCCTTGGAGCC 322
QY 309 CGCGCCGAGTGGTCTGCTACTCTGGGGTGCATCTCCAGGACGGGCCCTTGGAGCGGCG 368
Db 323 CGCGCCGAGTGGTCTGCTACTCTGGGGTGCATCTCCAGGACGGGCCCTTGGAGCGGCGC 382
QY 369 GCACACCGCGCAGTGGCGGCCATCGTGTGCGCGCCAACTACAGCAAGTGGAGCTGG 428
Db 383 GCACACCGCGCAGTGGCGGCCATCGTGTGCGCGCCAACTACAGCAAGTGGAGCTGG 442
QY 429 CGCGACCTGGCCCTCTGCGCTCTCAGTGTTCATGACGAATGGAGCCTTGGAGCC 488
Db 443 CGCGACCTGGCCCTCTGCGCTCTCAGTGTTCATGACGAATGGAGCCTTGGAGCC 502
QY 489 TGTCTGCTGCCCCCGCGCTCAGCAGCGCTTGTGTCAGCGCACCGCTGCTGGGCCACCGG 548
Db 503 TGTCTGCTGCCCCCGCGCTCAGCAGCGCTTGTGTCAGCGCACCGCTGCTGGGCCACCGG 562
QY 549 CTGGGAGACGCTCCA- GGAGGACAGATCTCTGCTCTCCCTGGGTGCTAC- AGGAAGTG 606
Db 563 CTGGGAGACGCTCCA- GGAGGAGACGATCTCTGCTCTCCCTGGGTGCTAC- AGGAAGTG 622
QY 607 GAGCTAAGGCTCTGGGCG- AGGCCACCTGTCAATGTCTTACAGCCA 653
Db 623 GAGCTAAGGCTCTGGGCGAAGGCAACTGGTCAATGTCTTACANNCA 670

RESULT 10
AX342936/c
LOCUS
DEFINITION
Sequence 3 from Patent WO0198467.
ACCESSION AX342936
VERSION AX342936.1 GI:18152214
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Xiao,Y. and Morozov,V.
Regulation of human prostatic-like serine protease
```


181 CCGGGACCTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 240
 Db GCGGGTCACTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 367
 QY 241 TCCTCTCATCGCCCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 300
 Db 368 TCCTCTCATCGCCCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 420
 QY 301 TTGGAGCCCGCGGCGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 360
 Db 421 --GCACACACAGGAGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 478
 QY 361 GAGCGCGCGACACCGCGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 420
 Db 479 GAGCGCGCGACACCGCGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 538
 QY 421 GAGCTGGGCGCGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 480
 Db 539 GGCTTGGGCGCGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 598
 QY 481 GTGTGGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 540
 Db 599 ATCGGGCGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 658
 QY 541 GCCACCGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 600
 Db 659 GTCACTGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 718
 QY 601 GAAGTGGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 660
 Db 719 CAATCGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 778
 QY 661 CCCTTCACTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 720
 Db 779 AAGCTTGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 838
 QY 721 GCGAGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 780
 Db 839 GCGAGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 898
 QY 781 TGGTTCAGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 840
 Db 899 TGGTTCAGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 958
 QY 841 GTTTCAGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 900
 Db 959 GTTTCAGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 1018
 QY 901 GTTTCAGGCTTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 924
 Db 1019 CCTCGTGGTGGCTTGGCTTGGCACTGAGCTGCTGCACATGAGGTGGCCACATCTGCGGGGCG 1042

BC001462 1800 bp mRNA linear PRI 04-OCT-2003
 LOCUS Homo sapiens protease, serine, 8 (prolactin), mRNA (cDNA clone
 DEFINITION MGC:2133 IMAGE:3138532), complete cds.
 ACCESSION BC001462
 VERSION BC001462.2 GI:33976244
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1800)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S., Iqbal, N.A., Peters, G.J.,
 Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Rulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahy, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 1800)
 Strausberg, R.
 Direct Submission
 Submitted (12-DEC-2000) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 On Aug 19, 2003 this sequence version replaced gi:12655206.
 Contact: MGC help desk
 Email: cgaps-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@hghri.nih.gov
 Akhtar, N., Ayèle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Hachighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., McGaspi, R.,
 Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Teurgelon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 4 Row: j Column: 3
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 21536453.
 Location/Qualifiers
 1. .1800
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:2133 IMAGE:3138532"
 /tissue_type="Placenta, choriocarcinoma"
 /clone_lib="NIH MGC_21"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 1. .1800
 /gene="PRSS8"
 /note="synonym: PROSTASIN"
 /db_xref="LocusID:5652"
 /db_xref="MIM:600823"
 176. .1207
 /codon_start=1
 /product="prolactin preproprotein"
 /protein_id="AAH01462.1"
 /db_xref="GI:12655207"
 /db_xref="LocusID:5652"
 /translation="MAQKGVLPQGLGAVALLYLILSLRSTGAGAEAPCCVAPQAR
 ITGSSAVAGQWPMQVSYTYGVHVCVGSGLVSEQWVLSARCFPFSEHHEAYEVKGA

FEATURES
 source

gene

CDS

HQLDSYSEDKAVSTLKDIIIPHSYLOQSGODIALQLSRPITFSYRPICTLPAANA
SFPNGLCHTCTGWHVAPSVLSLLPKPQLOEVLPIKSTCNCLNYIDAKPEBPHVQ
EDWQACAGVGGKDCAGDSGLPLSCDVEGLWYLTGIVSMGDACGNRPVYTLASS
YASHIQKVTELQPRVVPQESQPDNLCGSHLAFPSAPAGLLRPLFLPIGLALG
LLSPWLSEH
305..1018
/note="Tryp SPC; Region: Trypsin-like serine protease"
/db_xref="CDD:smart0020"

misc_feature

ORIGIN

Query Match 21.1%; Score 232.4; DB 9; Length 1800;
Best Local Similarity 55.0%; Pred. No. 3.1e-27;
Matches 508; Conservative 0; Mismatches 401; Indels 15; Gaps 2;

QY 1 GGGCCCTGTCTGGCCCATGCCCCAGAGGGGCTCTGGGGCTGGGCGAGTGGGGCT 60
DB |||||
QY 158 GGGCCCTGTCTGGCCCATGCCCCAGAGGGGCTCTGGGGCTGGGCGAGTGGGGCT 217
DB |||||
QY 61 GTGGCCAAATTCGACTCATCTCACTTTACGCGTTGGTGCCTGTCGAGCCGCTAGGGGC 120
DB |||||
QY 218 GTGGCCAAATTCGACTCATCTCACTTTACGCGTTGGTGCCTGTCGAGCCGCTAGGGGC 277
DB |||||
QY 121 CCCCCTGTAATCGGGGGCCCTGAGCCCTCGGCCCGCATCTGGGGGGCTCAAAAGCGCAG 180
DB |||||
QY 278 GCTCCCTGCGGTGTG-----GCCCCCAAGCAGCATCACAGGTGGCAGCAGTGCAGTC 331
DB |||||
QY 181 CCGGCACTCGCCCTTGGCAAGTGAAGCTGACCATGAGAGTGGCCACATCTGGGGGGGC 240
DB |||||
QY 332 GCGGTGAGTGGCCCTGACAGTTCAGCATCACTATGAGGGGTCCATGTGTGTGTGGTGGC 391
DB |||||
QY 241 TCCCTCAATCGCCCTCTGGTCTCTCCGCTGCTCACTGTTTCATGACGAATGGGAGC 300
DB |||||
QY 392 TCTCTCGGTCTGAGCAGTGGTGTGTGTCAGTGTCTCACTGTCTTCCCCAGCGA----- 444
DB |||||
QY 301 TTGAGCCCGGGCGAGTGGTGTGTAATGCTGGGCGGTGCACTCCAGGACGGGCCCTG 360
DB |||||
QY 445 --GCACCAAAAGGAGCCTATGAGGTCAAGTGGGGGGCCCAACAGCTAGACTCTCTACTCC 502
DB |||||
QY 361 GACGCGCGCACACCCGCGCAGTGGCCGCTCGTGGTGGCGGCACTACAGCCAAAGTG 420
DB |||||
QY 503 GAGNAGCCCAAGTTCAGCACCTTGAGGACATCATCCCCACCCAGCTACTCTCAGGAG 562
DB |||||
QY 421 GAGCTGGGCGCGACCTGGCCCTGTCGCTGCGCTCAACCGGCGAGCTGGGGCCCGCC 480
DB |||||
QY 563 GGTCCCAAGGCGCATTTGCACTCTCTCCAACTCAGCAGACCCATCACCTTCTCCCGTAC 622
DB |||||
QY 481 GTGTGGCTGTCTGCTGCGCGCCCTCACACCGCTTGTGTCAGGCGACCGCTGTGG 540
DB |||||
QY 623 ATCCGGCCCATCTGCTCCCTGACGCCAAGCGCTCTCTTCCCCAACGGGCTCCACTGCACT 682
DB |||||
QY 541 GCCACGGCTGGGAGAGCTCCAGGAGCGAGATCCTCTGCGCTCTCCCTGGGTGCTACAG 600
DB |||||
QY 683 GTCACTGCTGGGTCTATGTGGCCCTCAGTAGAGCTCTTGACGCCCAAGCCACTGAG 742
DB |||||
QY 601 GAAGTGGAGTAAAGCTGTGGGCGAGGCCACTGTCAATGTCTTACAGCAGCCCGGT 660
DB |||||
QY 743 CAACCTCGAGGTGCTCTGATCAGTGTGAGAGCTGTAACTGCTGCTGTAACTACACGCC 802
DB |||||
QY 661 CCCTTCAACTCACTCTCCAGATATTCAGGAGTGTGTGTGGTGTGTGTGTGTGTGTGT 720
DB |||||
QY 803 AAGCTGAGGAGCCGACCTTGTCCAAAGGAGACATGGTGTGTGTGTGTGTGTGTGTGT 862
DB |||||
QY 721 CCGAGGACACCTGCGAGGTGACTCTGGGGGGCCCTGGTGTGTGTGAGGAAGCGCGCCGC 780
DB |||||
QY 863 GGCNAGGAGCCCTGCCAGGCTGACTCTGGGGGGCCCACTCTCTGCGCTGTGGAGGTCTC 922
DB |||||
QY 781 TGGTTCCAGCGAATACAGCTTTGGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840
DB |||||
QY 923 TGGTACTGACGCGCAATTGTGAGCTGGGAGATGCTGTGGGGGGCCCGCAACAGCGCTGT 982
DB |||||
QY 841 GTTTTCACTGTGTGGTCACTATGAGGATGATGATACGGGAGCAGTGTGTGTGTGTGT 900
DB |||||
QY 983 GTGTACTCTGCGCTCCAGCTATGCTCTCTGGATCCAAAGCAAGGTGACAACTCCAG 1042
DB |||||

QY 901 CCTGGGCTGCTTTCACCCAG 924
DB 1043 CCTGCTGTGCTCCCAAAACCAG 1066

RESULT 15
AX098215 1835 bp DNA linear PAT 30-MAR-2001
LOCUS
DEFINITION Sequence 127 from Patent WO0118542.
ACCESSION AX098215
VERSION AX098215.1 GI:13515326
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Lee, J., Thompson, P. and Lillie, J.
TITLE Identification, assessment, prevention, and therapy of ovarian cancer
JOURNAL Patent: WO 0118542-A 127 15-MAR-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
source location/Qualifiers
1..1835
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 21.1%; Score 232.4; DB 6; Length 1835;
Best Local Similarity 55.0%; Pred. No. 3.1e-27;
Matches 508; Conservative 0; Mismatches 401; Indels 15; Gaps 2;

QY 1 GGGCCCTGTCTGGCCCATGCCCCAGAGGGGCTCTGGGGCTGGGCGAGTGGGGCT 60
DB 212 GGGCCCTGTCTGGCCCATGCCCCAGAGGGGCTCTGGGGCTGGGCGAGTGGGGCT 271
DB |||||
QY 61 GTGGCCAAATTCGACTCATCTCACTTTACGCGTTGGTGCCTGTCGAGCCGCTAGGGGC 120
DB |||||
QY 272 GTGGCCAAATTCGACTCATCTCACTTTACGCGTTGGTGCCTGTCGAGCCGCTAGGGGC 331
DB |||||
QY 121 CCCCCTGTAATCGGGGGCCCTGAGCCCTCGGCCCGCATCTGGGGGGCTCAAAAGCGCAG 180
DB 332 GCTCCCTGCGGTGTG-----GCCCCCAAGCAGCATCACAGGTGGCAGCAGTGCAGTC 385
DB |||||
QY 181 CCGGCACTCGCCCTTGGCAAGTGAAGCTGACCATGAGAGTGGCCACATCTGGGGGGGC 240
DB 386 GCGGTGAGTGGCCCTGAGAGTTCAGCATCACTATGAGGCGTCCATGTGTGTGTGGTGGC 445
DB |||||
QY 241 TCCCTCAATCGCCCTCTCTGGGTCTCTCTCGCTGCTCACTGTTTCAAGCAATGGGAGC 300
DB 446 TCTCTCGTGTCTGAGCAGTGGTGTCTGTCAGTGTCTCACTGCTTCCCCAGCGA----- 498
DB |||||
QY 301 TTGAGCCCGGGCGAGTGGTGTGTAATGCTGGGCGTGTGCTCCAGGACGGGCCCTG 360
DB 499 --GCACCAAAAGGAGCCTATGAGGTCAAGTGGGGGGCCCAACAGCTAGACTCTCTACTPCC 556
DB |||||
QY 361 GACGCGCGCACACCCGCGCAGTGGCCCATCTGTGTGGTGGCGGCAACTACAGCCAAAGTG 420
DB 557 GAGNAGCCCAAGTTCAGCACCTTGAGGACATCATCCCCACCCAGCTACTCTCAGGAG 616
DB |||||
QY 421 GAGCTGGGCGCGACCTGGCCCTCTGCGCTGCGCTTCAACCGGCGAGCCTGGGGCCCGCC 480
DB 617 GGTCCCAAGGCGCAGTTCAGTCACTCTTCCAACTCAGCAGACCCATCACCTTCTCCCGTAC 676
DB |||||
QY 481 GTGTGGCTGTCTGCTGCCCGGCGCTCACACCGCTTGTGTGCGAGGACCGCCCTGTGG 540
DB 677 ATCCGGCCCATCTGCTCTCTTGCAGCCAAACCGCTCTTCCCCAAACGGCTCCACTGCACT 736
DB |||||
QY 541 GCCACCGGCTGGGAGAGTCCAGGAGCAGATCTCTGCTGCTCTCTCTCTCTCTCTCTCTCT 600
DB 737 GTCACTGCTGGGTGTCATGTGGCCCCCTCTCAGTGAAGCTCTCTGACGCCCAAGCACTGAG 796
DB |||||

QY	601	GAAGTGGAGCTAAGGTGCTGGGCGAGGCCA	CCTGTCAATGTCCTACAGCCAGCCCGGT	660
Db	797	CAACTCGAGTGCCTCTGATCACTGCTGAGAC	CTGTAACTGCTGTACACATCGACGCC	856
QY	661	CCCTTCAACCTCACTCTCCAGATATTGCCAG	GGATGCTGTGTGCTGTGCTACCCAGAGGGC	720
Db	857	AAGCCTGAGGAGCCGCACTTTGTCCAAGAG	GCATGCTGTGTGCTGTGCTATGTGAGGGG	916
QY	721	CGCAGGGACACCTGCCAGGCTGACTCTGGG	GGCCCTGTGTGAGGAAGGCCGCCG	780
Db	917	GGCAGAGCCCTGCCAGGCTGACTCTGGG	GGCCCACTCTCCCTGTGGAGGCTTC	976
QY	781	TGGTTCAGGCAGGAATCACAGCTTTGGG	TTTGGCTGTGGACGGAGAAACCGCCCTGGA	840
Db	977	TGGTACCTGACGGGCATTGTGAGCTGGG	GAGATGCCCTGTGGGGCCCGCAACAGGCCTGT	1036
QY	841	GTTTTCACTGCTGTGGTACCTATGAGGC	ATGATACGGGAGCAGGTGATGGGTTGAG	900
Db	1037	GTGTACACTGTGGCCTCCAGCTATGCT	CTGATCCAAAGCAAGGTGACAGAACTCCAG	1096
QY	901	CCTGGGCTGCCCTTCCCAACCCAG	924	
Db	1097	CTCTGTGTGTGCCCAACCCAG	1120	

Search completed: February 25, 2004, 10:31:04
Job time : 4339 secs

WO200253742-A2.

07-JAN-2002: 2002WO-US000375.

05-JAN-2001: 2001US-0260018P.

08-JAN-2001; 2001US-0200500E:
28-FEB-2001: 2001US-0272411P:

02-MAR-2001; 2001US-0272817P.
 05-JUL-2001; 2001US-0303231P.
 12-JUL-2001; 2001US-0305060P.
 10-SEP-2001; 2001US-0318405P.
 12-SEP-2001; 2001US-0318700P.
 04-JAN-2002; 2002US-00037417.
 (CURA-) CURAGEN CORP.
 Kekuda R, Alsobrook JP, Tchernev VT, Liu X, Spytek KA;
 Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CAM, Li L;
 Goran L, Edinger S, Sciore P, Ellerman K, Malyankar U;
 Rothenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;
 Padigaru M, Taupier RJ, Miller CE, Eisen A;
 WPI; 2002-583619/62.
 P-PSDB; ABH09523.
 Novel polypeptides and nucleic acids homologous to transmembrane
 receptor, thymosin, neuromodulin-like family of proteins for diagnosing,
 treating cancer, atherosclerosis, neurological, skin and autoimmune
 disorders.
 Claim 9a; Page 142; 323pp; English.
 The invention relates to 24 novel human proteins designated NOV1-NOV14
 (ABB09501-ABB09524), collectively referred to as NOVX proteins, and
 nucleic acids encoding them (ABQ93875-ABQ93902). NOVX proteins and
 nucleotides are useful in the treatment, diagnosis or prevention of NOVX-
 associated disorders or in the manufacture of a medicament for treating
 such disorders, with specific applications described for each of the 24
 NOVX proteins, based on their homology to known proteins. Various
 disorders are associated with NOVX proteins including neurological
 disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
 pain, behavioural disorders, addiction, tuberculous sclerosis, cancers
 (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
 (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,
 various forms of arthritis, diabetes, thyroiditis, cardiovascular disease
 (e.g., hypertension), reproductive disorders, endometriosis,
 incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,
 cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
 disorders, obesity, bacterial infections and particularly cardiomyopathy,
 atherosclerosis, cell signal processing-related disorders and disorders
 of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
 be used to identify cellular receptors or downstream effectors which
 binds to a NOVX protein, and are also useful as targets for the
 identification of small molecules that modulate or inhibit processes such
 as neurogenesis, cell differentiation, cell motility, cellular
 proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
 nucleic acid sequences can be used to identify a cell or tissue type and
 are useful as a source of primers or probes for forensic biology and for
 identifying and cloning NOVX homologues in other cell types. Cells
 comprising NOVX nucleic acids are useful for producing non-human
 transgenic animals which are useful for studying the function and
 activity of NOVX proteins and for identifying and evaluating modulators
 of NOVX activity. The present sequence represents DNA encoding the
 prostatic precursor-like protein NOV14a. The gene encoding NOV14a is
 located on chromosome 16
 Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;
 Query Match 100.0%; Score 1102; DB 6; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 8.2e-185;
 Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGCTCTGGGGCTGGGCGAGCTGGGGGCT 60
 1 GGGCCCTTGTCTGGGCGCATGGGCGAGAGGGGCTCTGGGGCTGGGCGAGCTGGGGGCT 60
 61 GTGGCCAAATCTGACTCATCTACTCTTACGGCTTGGTGGCGCTGGGACCCGCTAGGGGC 120
 61 GTGGCCAAATCTGACTCATCTACTCTTACGGCTTGGTGGCGCTGGGACCCGCTAGGGGC 120

QY 121 CCCCCCTACTGGGGGCGCCCTGAGCCCTCGGCCCGCATCTGGGGGGGCTCAACGGCGAG 180
 Db 121 CCCCCCTACTGGGGGCGCCCTGAGCCCTCGGCCCGCATCTGGGGGGGCTCAACGGCGAG 180
 QY 181 CCGGGCACCTGGGCTTGGGCAAGTGAAGCTGACCTGACCATGGAGGTGGCCACATCTCGGGGGG 240
 Db 181 CCGGGCACCTGGGCTTGGGCAAGTGAAGCTGACCTGACCATGGAGGTGGCCACATCTCGGGGGG 240
 QY 241 TCCCTCATCGCCCCCTCTCTGGGTCTCTCGGTGCTCACTGTCTTATGAGCAATGGGACG 300
 Db 241 TCCCTCATCGCCCCCTCTCTGGGTCTCTCGGTGCTCACTGTCTTATGAGCAATGGGACG 300
 QY 301 TTGGAGCCCGCGCCGAGTGGTCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTG 360
 Db 301 TTGGAGCCCGCGCCGAGTGGTCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTG 360
 QY 361 GAGCGCGCGCACACCCGCGCAGTGGCGCCCATCTGGTGGCGGCCCACTACTACAGCAAGTG 420
 Db 361 GAGCGCGCGCACACCCGCGCAGTGGCGCCCATCTGGTGGCGGCCCACTACTACAGCAAGTG 420
 QY 421 GAGCTGGGCGCCGAGTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTG 480
 Db 421 GAGCTGGGCGCCGAGTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTG 480
 QY 481 GTGTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTG 540
 Db 481 GTGTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTGGGCTGCTGCTG 540
 QY 541 GCCACCGGCTGGGAGAGCTGCCAGGAGGAGATCTCTGCTGCTCTCTCTCTCTCTCTCTCTCAG 600
 Db 541 GCCACCGGCTGGGAGAGCTGCCAGGAGGAGATCTCTGCTGCTCTCTCTCTCTCTCTCTCTCAG 600
 QY 601 GAAGTGGAGCTTAAGCTGCTGGGCGAGCCCACTGCTCAATGCTCTCTACAGCCAGCCGCT 660
 Db 601 GAAGTGGAGCTTAAGCTGCTGGGCGAGCCCACTGCTCAATGCTCTCTACAGCCAGCCGCT 660
 QY 661 CCCTTCAACCTCACTCTCCAGATATTCGAGGAGTCTGTGTGCTGCTGCTGCTGCTGCTGCTG 720
 Db 661 CCCTTCAACCTCACTCTCCAGATATTCGAGGAGTCTGTGTGCTGCTGCTGCTGCTGCTGCTG 720
 QY 721 CGCAGGACACCTGCCAGGAGTCTCTGGGGGGGCTCTGTGTGAGGAGGAGGCGGCGC 780
 Db 721 CGCAGGAGACCTGCCAGGAGTCTCTGGGGGGGCTCTGTGTGAGGAGGAGGCGGCGC 780
 QY 781 TGGTTCAGGAGGAGATCAACAGCTTTGGGTTGGCTGTGAGCGGAGGAGAACCGCCCTGGA 840
 Db 781 TGGTTCAGGAGGAGATCAACAGCTTTGGGTTGGCTGTGAGCGGAGGAGAACCGCCCTGGA 840
 QY 841 GTTTTCACTGTGGCTTACCTATGAGGATGAGATCGGGAGGAGTGTGATGGGTTTCAGAG 900
 Db 841 GTTTTCACTGTGGCTTACCTATGAGGATGAGATCGGGAGGAGTGTGATGGGTTTCAGAG 900
 QY 901 CTTGGGCTGCTCTTCCACCCAGCCAGAGACCCAGTCAAGATGTTTATCATCAAAAG 960
 Db 901 CTTGGGCTGCTCTTCCACCCAGCCAGAGACCCAGTCAAGATGTTTATCATCAAAAG 960
 QY 961 GCATTCTGGATCTGACAGATCTTTTGGGCTTGGCTGTGAGGAGGAGTGTGATGAGTGTG 1020
 Db 961 GCATTCTGGATCTGACAGATCTTTTGGGCTTGGCTGTGAGGAGGAGTGTGATGAGTGTG 1020
 QY 1021 TCAACTGGGAGCCAAAGCTTGTCT 1080
 Db 1021 TCAACTGGGAGCCAAAGCTTGTCT 1080
 QY 1081 TGGGGGTTCTGATGGGGCTCTCC 1102
 Db 1081 TGGGGGTTCTGATGGGGCTCTCC 1102

RESULT 2
 ABQ93902
 ID ABQ93902 standard; DNA; 1102 BP.
 XX

QY 601 GAGTGGAGCTAAGCTGCTGGGGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGT 660
Db |||||
QY 601 GAGTGGAGCTAAGCTGCTGGGGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGT 660
Db |||||
QY 661 CCCTTCAACCTCACTCTCCAGATATTCAGGGAGTGTGTGTGCTACCCAGAGGGC 720
Db |||||
QY 661 CCCTTCAACCTCACTCTCCAGATATTCAGGGAGTGTGTGTGCTACCCAGAGGGC 720
Db |||||
QY 721 CGCAGGACACCTGCCAGGGTGAATCTGGGGGGCCCTGGTCTGTGAGGAAGGGCGCGC 780
Db |||||
QY 721 CGCAGGACACCTGCCAGGGTGAATCTGGGGGGCCCTGGTCTGTGAGGAAGGGCGCGC 780
Db |||||
QY 781 TGGTTCACAGGAGGAATCACCAGCTTTTGGCTGTGGACGGAGAAACCGCCCTGGA 840
Db |||||
QY 781 TGGTTCACAGGAGGAATCACCAGCTTTTGGCTGTGGACGGAGAAACCGCCCTGGA 840
Db |||||
QY 841 GTTTCATCTGTGGCTACCTATGAGGCATGGATACGGAGCAGGTGATGGTTCAGAG 900
Db |||||
QY 841 GTTTCATCTGTGGCTACCTATGAGGCATGGATACGGAGCAGGTGATGGTTCAGAG 900
Db |||||
QY 901 CCTGGGCTGCTTTCCACACCCAGCCAGAGACCCAGTCAGATTGTTTACATCAAACG 960
Db |||||
QY 901 CCTGGGCTGCTTTCCACACCCAGCCAGAGACCCAGTCAGATTGTTTACATCAAACG 960
Db |||||
QY 961 GCATTCTGGATCTGCCAGATCCTTTTGGAGCCCTTGTCCCATATATCAGTAGAGTC 1020
Db |||||
QY 961 GCATTCTGGATCTGCCAGATCCTTTTGGAGCCCTTGTCCCATATATCAGTAGAGTC 1020
Db |||||
QY 1021 TCAACTGGGACCAAAAGCTTGTCTCCCTGGCTCTCTCCACACTCTCTCTGGGCTC 1080
Db |||||
QY 1021 TCAACTGGGACCAAAAGCTTGTCTCCCTGGCTCTCTCCACACTCTCTCTGGGCTC 1080
Db |||||
QY 1081 TGGGGGTTCTGATGGGGCTCC 1102
Db |||||
QY 1081 TGGGGGTTCTGATGGGGCTCC 1102
Db |||||

RESULT 3
ID ABK31795
XX ABK31795 standard; DNA; 2457 BP.
AC ABK31795;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA encoding novel human protease #52.
XX
KW Human; protease; cancer; immune-related disorder; cardiovascular disease;
KW neuronal-associated disease; metabolic disorder; inflammatory disorder;
KW nervous system disorder; sexual dysfunction; pain; mood disorder;
KW hypertension; psychotic disorder; neurological disorder; dyskinesia;
KW viral infection; human immunodeficiency virus; HIV; non-viral infection;
KW ocular disease; cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200860-A2.
XX
PD 03-JAN-2002.
XX
PF 26-JUN-2001; 2001WO-US020171.
XX
PR 26-JUN-2000; 2000US-0214047P.
XX
XX (SUGEN-) SUGEN INC.
XX
XX Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
PI Charyczak G;
XX
XX WPI; 2002-139913/18.
XX
XX P-PSDB; AAU82753.
XX
XX Nucleic acids encoding novel human proteases, useful for use for

PT treating diseases and disorders such as cancers, immune-related diseases
PT and disorders, cardiovascular disease (e.g. restenosis) and inflammatory
PT disorders.
XX Claim 26; Fig 1SS-TT; 313pp; English.
XX
XX The present invention relates to the isolation of novel human proteases,
CC and the nucleic acids encoding them. The sequences of the invention are
CC useful for treating diseases and disorders such as cancers (e.g. breast,
CC colon, lung), immune-related diseases and disorders (e.g. inflammatory
CC diseases and asthma), cardiovascular diseases (e.g. restenosis and
CC coronary thrombosis), brain or neuronal-associated diseases, metabolic
CC disorders (e.g. diabetes, obesity), inflammatory disorders (e.g.
CC rheumatoid arthritis and psoriasis), central or peripheral nervous system
CC diseases, migraines, pain, sexual dysfunction, mood disorders, attention
CC disorders, neurological disorders, hypotension, hypertension, psychotic
CC disorders, cognitive disorders (e.g. Alzheimer's disease, Parkinson's
CC disease) and dyskinesias. The nucleic acids and polypeptides are also
CC useful for treating viral infections caused by human immunodeficiency
CC virus (HIV), and non-viral infections such as ocular disease (e.g.
CC glaucoma) and macular degeneration. ABK31744-ABK31802 represent DNA
CC sequences encoding for the novel human proteases of the invention
XX
SQ Sequence 2457 BP; 377 A; 851 C; 797 G; 432 T; 0 U; 0 Other;

Query Match 74.0%; Score 815.6; DB 6; Length 2457;
Best Local Similarity 99.5%; Pred. No. 1.6e-134;
Matches 818; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 124 CCGTACTGCGGGCCCTGAGCCCTCGCGCCGATCGTGGGGGGCTCAAAAGCGCAGCGC 183
Db |||||
QY 106 CTGAGCTGCGGGGCCCTGAGCCCTCGCGCCGATCGTGGGGGGCTCAAAAGCGCAGCGC 165
Db |||||
QY 184 GGCACCTCTGGCAAGTGAGCTGACCATGAGAGTGGCCACATCTGGGGGGCTCC 243
Db |||||
QY 166 GGCACCTGGCAAGTGAGCTGACCATGAGAGTGGCCACATCTGGGGGGCTCC 225
Db |||||
QY 244 CTGATGCGCCCTCTCTGGGTCTCTCGCTGCTCACTGTTTCATGACGAATGGGACGCTG 303
Db |||||
QY 226 CTGATGCGCCCTCTCTGGGTCTCTCGCTGCTCACTGTTTCATGACGAATGGGACGCTG 285
Db |||||
QY 304 GAGCCCGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363
Db |||||
QY 286 GAGCCCGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 345
Db |||||
QY 364 GCGCGCACACCCCGCGAGTGGCCGCTATCGTGGTGGCGGCAACTACAGCAAGTGGAG 423
Db |||||
QY 346 GCGCGCACACCCCGCGAGTGGCCGCTATCGTGGTGGCGGCAACTACAGCAAGTGGAG 405
Db |||||
QY 424 CTGGCGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
Db |||||
QY 406 CTGGCGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 465
Db |||||
QY 484 TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543
Db |||||
QY 466 TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
Db |||||
QY 544 ACCGCTGGGAGAGCTCCAGAGCAGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 603
Db |||||
QY 526 ACCGCTGGGAGAGCTCCAGAGCAGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 585
Db |||||
QY 604 GTGGAGCTAAGGCTGCTGGGGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGCTCC 663
Db |||||
QY 586 GTGGAGCTAAGGCTGCTGGGGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGCTCC 645
Db |||||
QY 664 TTCAACCTCACTCTCCAGATATTCAGGGAGTGTGTGTGCTGCTGCTGCTGCTGCTGCTG 723
Db |||||
QY 646 TTCAACCTCACTCTCCAGATATTCAGGGAGTGTGTGTGCTGCTGCTGCTGCTGCTGCTG 705
Db |||||
QY 724 AGGGACACTGCGAGGGTGAATCTCTGGGGGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 783
Db |||||
QY 706 AGGGACACTGCGAGGGTGAATCTCTGGGGGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765
Db |||||

784	QY	TTCCAGGGCAGGATACACAGCTTTGGGTTTGGCTCTGCGACGGAGAAACCGCCCTCGAGTT	843
766	Db	TTCCAGGCAGGAAATCACAGCTTTGGCTTTGGCTGTGGACGGAGAAACCGCCCTCGAGTT	825
844	QY	TTCACTGCTGTGGTCTACTATGAGGCATGGATACGGGACAGGTGATGGTTTCAGAGCTT	903
826	Db	TTCACTGCTGTGGTCTACTATGAGGCATGGATACGGGACAGGTGATGGTTTCAGAGCTT	885
904	QY	GGGCTGTGCTTTTCCACCCAGCCCGCAGAAACCCAGTCAGAT	945
886	Db	GGGCTGTGCTTTTCCACCCAGCCCGCAGAAACCCAGTCAGAT	927

RESULT 4
ABQ75956
ID ABQ75956 standard: cDNA: 2681 BP.

AC ABQ75956;

17-OCT-2002 (first entry)

DE Human PMMM encoding sequence Incyte ID 2751509CBI.

Human; PMMV; protein modification and maintenance molecule;
anticoagulant; neuroprotective; neotropic; cytosolic; antipsoriatic;
antialthmatic; dermatological; antidiabetic; antiparkinsonian;
antianemic; antiinflammatory; antiulcer; antianginal; cardiant;
hepatotropic; osteopathic; antiemetic; antipyretic; virucide;
antibacterial; fungicide; gastrointestinal; antidiarrheic; laxative;
haemostatic; anti-HIV; thrombolytic; anticoagulant; gynaecological;
cardiovascular; antiarteriosclerotic; hypotensive; vasotropic;
antitumour; antirheumatic; immunosuppressive; antiallergic; antithyroid;
nephrotropic; antigout; thymimetic; aniathrictic; uropathic;
ophthalmological; antiparasitic; tranquiliser; vulnerary; keratolytic;
auditory; antiseborrheic; antidepressant; neuroleptic; antinfertility;
anthelmintic; protozoacide; Crohn's disease; hypertension; autoimmune;
inflammatory; anaemia; cell proliferative; developmental; epithelial;
scabies; neurological; Alzheimer's disease; reproductive;
ectopic pregnancy; gene therapy; vaccine; disorder; prostatic; gene; ss.

OS Homo sapiens.

Key	Location/Qualifiers
1A	
FH	

ET	CDS	1.2457
----	-----	--------

```
FT /*tag=
```

FT /product= "protein modification and maintenance molecule"

PN WO200246383-A2.

13-JUN-2002.

05-DEC-2001: 2001WO-US046964.

08-DEC-2000: 2000US-0254399P.

PR 21-DEC-2000; 2000US-0237803F.
PR 05-JAN-2001; 2001US-0260110P

PR 19-JAN-2001; 2001US-0262851P.
PB 25-JAN-2001; 2001US-0354533P.

[illegible]

PI Lal PG, Walia NK, Hafalia AJA, Gandhi AR, Au-Young J, Elliott VS;

PI Ramkumar J, Thangavelu K, Lu Y, Warren BA, Lu DAM, Lee EA;

PI Tribouley CM, Arvizu C, Delegeane AM, Yao MG, Khan FA;

XX

DR WPI; 2002-519664/55.

XX
XX

PT New isolated Protein Modification and Maintenance polypeptides,

PT for diagnosis, and treatment of e.g. gastrointestinal disorders.

PS The invention relates to an isolated Protein Modification and Maintenance
XX (PMM) polypeptide. Polypeptides of the invention may be used in the
CC diagnosis, treatment and prevention of disorders associated with
CC decreased expression or activity of PMM. These include gastrointestinal
CC disorders (e.g. Crohn's disease), cardiovascular disorders (e.g.
CC hypertension), autoimmune/inflammatory disorders (e.g. anaemia), cell
CC proliferative disorders, developmental disorders, epithelial disorders
CC (e.g. scabies), neurological disorders (e.g. Alzheimer's disease)
CC reproductive disorders (e.g. ectopic pregnancy), and in gene therapy or a
CC vaccine for such diseases. They may also be used in the assessment of the
CC effects of exogenous compound on the expression of nucleic acid and amino
CC acid sequences of protein modification and maintenance molecules. The
CC current sequence represents a human PMM encoding sequence of the
CC invention, encoding a polypeptide which has been found to have homology
CC with rat prostaasin
XX Sequence 2681 BP: 431 A: 919 C: 861 G: 470 T: 0 U: 0 Other:

Query Match	74.0%;	Score 815.6;	DB 6;	Length 2681;
Best Local Similarity	99.5%;	Pred. No. 1.6e-134;		
Matches 818;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	124	CGGACTCGGGGCGCCCTGAGCCCTCGGCGCATCGTGGGGGGCTCAAAAGCGCGAGCGG	183	
Db	106	CTGGACTCGGGGCGCCCTGAGCCCTCGGCGCGCATCGTGGGGGGCTCAAAAGCGCGAGCGG	165	
QY	184	GGCACTCGCCCTTGCGAAGTGAGCGCTGCACCATGGAGGTGGCCACATCTCGGGGGCTCC	243	
Db	166	GGCACTCGCCCTTGCGAAGTGAGCGCTGCACCATGGAGGTGGCCACATCTCGGGGGCTCC	225	
QY	244	CTCATCGCCCGCTCTCGGTCTCTCGCGTGTCACTGTTCATGACGAATGGAGCGTGG	303	
Db	226	CTCATCGCCCGCTCTCGGTCTCTCGCGTGTCACTGTTCATGACGAATGGAGCGCTGG	285	
QY	304	GAGCCGCGCGCGAGTGTGCTGCTGCGGCTGCATCTCCAGGACGGGCGCCCTGGAC	363	
Db	286	GAGCCGCGCGCGAGTGTGCTGCTGCGGCTGCATCTCCAGGACGGGCGCCCTGGAC	345	
QY	364	GGCGGCGCACCCCGCGCAGTGGCGCCCATCGTGTGCGCGCCAACTACAGCCCAAGTGGAG	423	
Db	346	GGCGGCGCACCCCGCGCAGTGGCGCCCATCGTGTGCGCGCCAACTACAGCCCAAGTGGAG	405	
QY	424	CTGGGCGGCGACCTGGGCCCTTCTGGCGCTTGCCTCACCGCGCAGCTGGGCGCCCGCGTG	483	
Db	406	CTGGGCGGCGACCTGGGCCCTTCTGGCGCTTGCCTCACCGCGCAGCTGGGCGCCCGCGTG	465	
QY	484	TGGCGCTGTCTGCTGCTCCCGCGCGCTTCACACCGCTTCGTGCAAGGACCGCGCTGTGGGCGC	543	
Db	466	TGGCGCTGTCTGCTGCTCCCGCGCGCTTCACACCGCTTCGTGCAAGGACCGCGCTGTGGGCGC	525	
QY	544	ACCGGTGGGGAGACGTCCAGAGGACGATCCTCTGCGCTTCGCCCTGGGTGCTACAGGAA	603	
Db	526	ACCGGTGGGGAGACGTCCAGAGGACGATCCTCTGCGCTTCGCCCTGGGTGCTACAGGAA	585	
QY	604	GTGGAGCTTAAGCTGCTGGGCGAGGCCACCTGTCAATGTCTCTACGACGACCGCGTCC	663	
Db	586	GTGGAGCTTAAGCTGCTGGGCGAGGCCACCTGTCAATGTCTCTACGACGACCGCGTCC	645	
QY	664	TTCAACCTCACTCTCCAGATATTGCCAGGATGTGTGTGTGCTGCTACCCAGAGGGCGCG	723	
Db	646	TTCAACCTCACTCTCCAGATATTGCCAGGATGTGTGTGTGCTGCTACCCAGAGGGCGCG	705	
QY	724	AGGGAACCTGCGCAGGGTGAATCTGGGGGGGCCCTTGCTGTGTGAGGAAGGGCGCGCTGG	783	
Db	706	AGGGAACCTGCGCAGGGTGAATCTGGGGGGGCCCTTGCTGTGTGAGGAAGGGCGCGCTGG	765	
QY	784	TTCCAGGACAGGAATCACAGCTTTTGGGTTTGGCTGTGGACGGAGAAACCGCCTTGAGTT	843	
Db	766	TTCCAGGACAGGAATCACAGCTTTTGGCTGTGGACGGAGAAACCGCCTTGAGTT	825	
QY	844	TTCACTGTGTGGCTACCTATGAGGCATGGGATACGGGAGCAGGTGATGGGTTACAGACCT	903	

826 TTCACTGCTGGGTACCTATGAGCATGGATCGGAGCAGGTGAGGTTCAGAGCCT 885
 904 GGGGCTGCTTTCCACCCAGCCCAAGAACCCAGTCAGAT 945
 886 GGGGCTGCTTTCCACCCAGCCCAAGAACCCAGTCAGAT 927

RESULT 5
 ID AAK94500 standard; cDNA; 2810 BP.
 AC AAK94500;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human full-length cDNA, SEQ ID NO: 3346.
 XX
 XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 KW
 XX Homo sapiens.
 OS
 XX EP1130094-A2.
 FN
 XX 05-SEP-2001.
 PD
 XX 07-JUL-2000; 2000EP-00114089.
 PF
 XX 08-JUL-1999; 99JP-00194486.
 PR
 XX 11-JAN-2000; 2000JP-00118774.
 PR
 XX 02-MAY-2000; 2000JP-00183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 DR P-PSDB; AAM93568.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 XX
 PS Claim 8; SEQ ID NO 3346; 1380pp + Sequence Listing; English.
 CC
 CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a full length human cDNA of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in CD-ROM format directly
 CC from EPO
 XX
 SQ Sequence 2810 BP; 436 A; 964 C; 902 G; 508 T; 0 U; 0 Other;

Query Match 73.2%; Score 806.6; DB 4; Length 2810;
 Best Local Similarity 92.0%; Pred. No. 6.2e-133;
 Matches 862; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 10 TCCTGGGCGCATGCCAGAGGGGTTCCTGGGCGCTGGGCGCTGGGGGTGGGCCAAT 69
 DB 23 TCATAGCGCCATGCCCGGACCTGCTCTCCCTTGTGATGTTGTATCATGCTCCATC 82
 QY 70 TCCTGACTCATPACTCACTTTACGGGTTGGTCCGCTCCGAGACCCGCTAGGGGCCCGGTAC 129
 DB 83 CCAGGAGCCTTCAGGACTCAGTCTCAGTCCCTACCCAGGAACAACCTGAAGATCTGGAC 142
 QY 130 TGGCGGGCGCCTGAGCCCTCGGCCCGGATCGT-GGGGGGGCTCAACGCCGCGCGGGCAC 188

143 TGGGGGGCCCTGAGCCCTCGGCCCGCATGTTGGGGGGGCTCAAAAGCGCAGCGGGGCAC 202
 QY 189 CTGGCCTTGGCAAGTAGTGGCTGACCATGAGAGTGGCCACATCTGGGGGGCTCCCTCAT 248
 DB 203 CTGGCCTTGGCAAGTAGTGGCTGACCATGAGAGTGGCCACATCTGGGGGGCTCCCTCAT 262
 QY 249 CGCCCCCTCTGGGTCTCTCCGCTGCTCACTGCTTTCATGACGAATGGAGCTTTGGAGCC 308
 DB 263 CGCCCCCTCTGGGTCTCTCCGCTGCTCACTGCTTTCATGACGAATGGAGCTTTGGAGCC 322
 QY 309 CGGGGCCAGTGGTGGTACTGCTGGGCGTGCATCCAGGACGGGGCCCTTGGAGCGCGC 368
 DB 323 CGGGGCCAGTGGTGGTACTGCTGGGCGTGCATCCAGGACGGGGCCCTTGGAGCGCGC 382
 QY 369 GCACACCGCGCAGTGGCGCCCATCGGTGGTGGCGCCAACTACAGCCAAAGTGGAGCTGGG 428
 DB 383 GCACACCGCGCAGTGGCGCCCATCGGTGGTGGCGCCAACTACAGCCAAAGTGGAGCTGGG 442
 QY 429 CGCGACCTGGCCCTGCTGGGCGCTTCAACCGCCAGCGCTGGGGCCCGCGCTGTGGCC 488
 DB 443 CGCGACCTGGCCCTGCTGGGCGCTTCAACCGCCAGCGCTGGGGCCCGCGCTGTGGCC 502
 QY 489 TGTGTGCTGGCCCGCGCCCTCAACCGCTTGTGTCACGGCACCGCCCTGCTGGGGCCACGG 548
 DB 503 TGTGTGCTGGCCCGCGCCCTCAACCGCTTGTGTCACGGCACCGCCCTGCTGGGGCCACGG 562
 QY 549 CTGGGGAGAGCTCCAGGAGGAGATCTCTGCTGCTCTCCCTGGGCTGTCTACAGGAAGTGG 608
 DB 563 CTGGGGAGAGCTCCAGGAGGAGATCTCTGCTGCTCTCCCTGGGCTGTCTACAGGAAGTGG 622
 QY 609 GCTAAGCTGTGGGCGAGGCCACTGTCTAATGTCTCTACAGCCAGCCCGCTTCCCTCAA 668
 DB 623 GCTAAGCTGTGGGCGAGGCCACTGTCTAATGTCTCTACAGCCAGCCCGCTTCCCTCAA 682
 QY 669 CTTCACTTCCAGATATTCAGGAGTGTCTGTGCTGGCTACCCAGAGGCGCGCAGGGA 728
 DB 683 CTTCACTTCCAGATATTCAGGAGTGTCTGTGCTGGCTACCCAGAGGCGCGCAGGGA 742
 QY 729 CACTTCCAGAGGTCACCTCTGGGGGGCCCTCGTGTGTGAGGAAGCGCGCTGGTTCCA 788
 DB 743 CACTTCCAGAGGTCACCTCTGGGGGGCCCTCGTGTGTGAGGAAGCGCGCTGGTTCCA 802
 QY 789 GGCAGGAATACACAGCTTGGGTTGGCTGTGAGCGGAGAAACCGCCCTTGAGTTTAC 848
 DB 803 GGCAGGAATACACAGCTTGGGTTGGCTGTGAGCGGAGAAACCGCCCTTGAGTTTAC 862
 QY 849 TGCTGTGGCTACCTATGAGGATGATACGAGGAGCAGGTGATGGTTTCAGAGCCTTGGGCC 908
 DB 863 TGCTGTGGCTACCTATGAGGATGATACGAGGAGCAGGTGATGGTTTCAGAGCCTTGGGCC 922
 QY 909 TGCCTTTTCCACCCAGCCCGGAGAACCCAGTTCAGAT 945
 DB 923 TGCCTTTTCCACCCAGCCCGGAGAACCCAGTTCAGAT 959

RESULT 6
 ABK13565/c
 ID ABK13565 standard; cDNA; 537 BP.
 XX
 AC ABK13565;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human prostatic-like serine protease cDNA #1.
 XX
 KW Human; prostatic-like serine protease; cytosolic; antiatherosclerotic;
 KW virucide; osteopathic; antinflammatory; vasotropic; neuroprotective;
 KW trypsin-like; mekasin; autoimmune lesion; atherosclerosis; gene;
 KW renal failure; malignant cell; tumour angiogenesis; enzyme; inflammation;
 KW chronic obstructive pulmonary disease; COPD; restenosis; atherosclerosis;
 KW neurodegenerative disease; prion protein; infection; amyloid plaque;
 KW Genstmann-Straussler Syndrome; viral infection; Scrapie;

Creutzfeldt-Jakob disease; metastatic cancer; lipid accumulation; osteoporosis; Paget's disease; ss.

Homo sapiens.

Key Location/Qualifiers
FH complement(3. .536)
FT /*tag= a
ET /product= "Proteinase-like enzyme"
ET /partial
ET /note= "No start or stop codons shown"

WO200198467-A2.

27-DEC-2001.

22-JUN-2001; 2001WO-EP007117.

23-JUN-2001; 2000US-0213588P.

20-MAR-2001; 2001US-0276909P.

(FARB) BAYER AG.

Xiao Y, Morozov V;

WPI; 2002-114576/15.

P-PSDB; AAU75082.

Novel human proteinase-like serine protease polypeptide and polynucleotide which can be regulated for treating metastasis of malignant cells, inflammation, atherosclerosis, neurodegenerative disease and infections.

Claim 1; Fig 1; 111pp; English.

This invention comprises the cDNA and protein sequences of an isolated proteinase-like serine protease and reagents and methods for regulating the human proteinase-like enzyme activity. Proteinase is a trypsin-like serine protease purified from human seminal fluid. An antibody specific for proteinase-like serine protease is useful for immunodetection and diagnosis of micro-metastases, autoimmune lesions and renal failure in biopsy specimens, plasma samples and body fluids. The antibody may be used to modulate enzyme activity in a disease, such as metastasis of malignant cells, tumour angiogenesis, inflammation, chronic obstructive pulmonary disease (COPD), atherosclerosis, neurodegenerative disease (e.g. prion protein amyloid plaques of Gerstmann-Strausler Syndrome, Creutzfeldt-Jakob disease, Scrapie) or infection, particularly viral infection. The human proteinase-like serine protease gene provides a therapeutic target of decreasing human proteinase-like serine protease activity, in particular for treating or preventing metastatic cancer. The agonists and antagonists of the nucleotide sequence may be used to mimic, augment and inhibit the enzyme activity which may be useful to treat osteoporosis, Paget's disease and degradation of bone implants particularly dental implants. Altered levels of human proteinase-like serine protease activity inhibits both smooth muscle cell proliferation and lipid accumulation and inhibit the progression of restenosis and atherosclerosis. The nucleic acid sequence is also useful in diagnostic assays for detecting diseases and abnormalities or susceptibility to diseases related to the presence of mutations in nucleic acid sequences which encode the enzyme. The present sequence represents the human proteinase-like serine protease #1 nucleotide sequence of the invention

Sequence 537 BP; 92 A; 176 C; 197 G; 72 T; 0 U; 0 Other;

Query Match 48.6%; Score 535.4; DB 6; Length 537;
Best Local Similarity 99.8%; Pred. No. 3e-85;
Matches 536; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

243 CCTCATCGCCCTCTCGGTCTCTCCGCTGCTCACTGTTTCATGACGAATGGAGCGTT 302
|||||
537 CCTCATCGCCCTCTCTCGGTCTCTCCGCTGCTCACTGTTTCATGACGAATGGAGCGTT 478
|||||
303 GGAGCCCGCGCGCGAGTGTGCGTACTGCTGGCGGTGCACTCCAGAGCGGCGCCCTGGA 362
|||||

Db 477 GGAGCCCGCGCGCGAGTGTGCTGCTGGCGGTGCTACCTCCAGAGCGGCGCCCTGGA 418
QY 363 CGGCGCGCACACCGCGCAGTGGCGGCATGCTGGTGCAGGCGCACTACAGCCAAAGTGA 422
Db 417 CGGCGCGCACACCGCGCAGTGGCGGCATGCTGGTGCAGGCGCACTACAGCCAAAGTGA 358
QY 423 GCTGGGCGCGCGACCTGCGCCCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 482
Db 357 GCTGGGCGCGCGACCTGCGCCCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 298
QY 483 GTGCGCTGCTGCGCTGCGCCCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 542
Db 297 GTGCGCTGCTGCGCTGCGCCCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 238
QY 543 CACCGCTGGGCGAGACCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 602
Db 237 CACCGCTGGGCGAGACCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 178
QY 603 AGTGGAGCTAAGGCTGCTGGGCGAGGCGACCTGTCATGTCATGTCATGTCATGTCATGTCATGTC 662
Db 177 AGTGGAGCTAAGGCTGCTGGGCGAGGCGACCTGTCATGTCATGTCATGTCATGTCATGTCATGTC 118
QY 663 CTTCAACCTCACTCTCCAGATATTTGCCAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 722
Db 117 CTTCAACCTCACTCTCCAGATATTTGCCAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 58
QY 723 CAGGACACCTGCGCAGGCTGCTGCTGGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779
Db 57 CAGGACACCTGCGCAGGCTGCTGCTGGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1

RESULT 7

AAK92190

ID AAK92190 standard; cDNA; 670 BP.

XX AAK92190;

DT 06-NOV-2001 (first entry)

DE Human cDNA 5'-end sequence, SEQ ID NO: 650.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

OS Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-00114089.

XX 08-JUL-1999; 99JP-00194486.

XX 11-JAN-2000; 2000JP-00118774.

XX 02-MAY-2000; 2000JP-00183765.

XX (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

XX 830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation.

XX Claim 2; SEQ ID NO 650; 1380pp + Sequence Listing; English.

XX The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5' and 3' ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA


```
Db 308 CCCGTCAGTGGCCCTGCGAGTGCAGATCACTATGAAGGGTCCATGTGTGTGGC 367
Qy 241 TCCCTCATCGCCCTCTGGGTCTCTCCGCTGTCTACTGTTCATGACGAATGGAGC 300
Db 368 TCTCTGTGTCTGAGCAGTGGGTGTGTGAGCTGTCTACTGTCTCCACAGGA----- 420
Qy 301 TTGGAGCCGGCCGAGTGGTCTGCTACTGTCTGGGCTGCACTCCACAGAGGGGCCCTG 360
Db 421 --GCACCAAGGAAGCCATAGAGTCAAGCTTGGGGGCCACACAGTACTCTACTCC 478
Qy 361 GACGGCGGCAACACCGGGCAGTGGCGGCCATCTGTGGTGGTGGCCCAACTACAGCCAAAGT 420
Db 479 GAGGAGCCCAAGTTCAGCACCTCTGAAGGACATCATCCCCACCCAGCTACTCTCCAGAG 538
Qy 421 GAGCTGGGGCCGAGCTGGCCCTGTGCGCTGGGCTCACCGCCAGCTGGGCCCCGCC 480
Db 539 GGCCTCCAGGGGCAATTCGACTCTCTCAACTTCAGCAGACCCATCACTTCTCCCGTAC 598
Qy 481 GTGTGGCTGTCTGCTGCTGCCCGGCTCACACCGCTTCTGTCAGGACAGCCCTGTCTGG 540
Db 599 ATCCGGCCCATCTGCTCTCCCTGACGCAAGGCTCTCTCCCAAGGCTCTCACHGCACT 658
Qy 541 GCCACGGCTGGGAGACCTCCAGGAGGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
Db 659 GTCACTGGCTGGGTCTATGTGGCCCTCAGTGAGCTCTCTGACGCGCCAAAGCCTGCGAG 718
Qy 601 GAACTGGAGTAAAGTCTGTGGCGGAGGCGCACTGTCAATGTCTCTACAGCAGCCCGGT 660
Db 719 CAACTCGAGGTGCTCTGTATCATGTCTGTGAGACGCTGTAACTGTCTGTATCAACATCAG 778
Qy 661 CCCTTCACTCTCTCCAGATATTGCGAGGATGCTGTGTGCTGGCTACCCAGAGGCG 720
Db 779 AAGCTGAGGAGCGGACATTTGTCAGAGGACATGTGTGTGTGTGTGTGTGTGTGTGTGT 838
Qy 721 CGCAGGGAACCTCCAGGGTCACTCTGGGGGGCCCCCTGTGTGTGTGTGTGTGTGTGTGTGT 780
Db 839 GSCAAGGAGCCTGCCAGGTGACTCTGGGGGCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 898
Qy 781 TGGTTCAGGAGGAATACCAAGCTTTGGCTGTGGAGTGTGGAGGAGAAACCGCCCTGGA 840
Db 899 TGGTACCTGACGGGCAATGTGAGCTGGGGAGATGCTGTGGGGCCCGCAACAGGGCTGT 958
Qy 841 GTTTTCACCTGTGTGCTACCTATGAGGCAATGATACGGGAGCAGGTGATGGTTTCAGAG 900
Db 959 GTGTACCTCTGCGCTTCCAGCTATGCTCTCTGATCCAAAGCAAGTGTACAGAACTCCAG 1018
Qy 901 CTGGGCTGCTTCCCAACCCAG 924
Db 1019 CCTCGTGTGTGGTGGCCCAACCCAG 1042

RESULT 12
AAF98720
ID AAF98720 standard; DNA; 1835 BP.
XX
AC AAF98720;
XX
DT 02-JUL-2001 (first entry)
DE
DE Human late stage ovarian tumour polynucleotide marker 28.
XX
XX Human; ovarian cancer; identification; detection; characterisation;
XX tumour; kinase; marker; cytostatic; antisense gene therapy; ds.
XX
OS Homo sapiens.
XX
XX WO200118542-A2.
XX
XX 15-MAR-2001.
XX
XX 01-SEP-2000; 2000WO-US024199.
XX
```

```
PR 03-SEP-1999; 99US-0152547P.
PR 16-MAR-2000; 2000US-0190347P.
PR 21-MAR-2000; 2000US-0191321P.
PR 31-MAY-2000; 2000US-0208382P.
PR 20-JUL-2000; 2000US-00220467.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX Lee J, Thompsho P, Lillie J;
XX WPI; 2001-211428/21.
DR
PT Detection, assessment, prevention and therapy of ovarian cancer,
PT comprises detecting changes in the expression of a variety of markers.
XX
XX Claim 1; Page 1186-1187; 1198pp; English.
CC
CC The present invention describes a method for assessing whether a patient
CC is afflicted with ovarian cancer by comparing: (1) the expression of a
CC marker (I) (see AAF98594 to AAF98730), in a patient sample; and (2) the
CC normal level of expression of (I) in a control non-ovarian cancer sample,
CC where a significant difference between the level of expression in (a) and
CC (b) is an indication that the patient is afflicted with ovarian cancer.
CC (I) have cytostatic activities and can be used in antisense gene therapy.
CC The method, compositions and kits from the present invention can be used
CC for: (1) assessing and treating ovarian cancer; (2) making isolated
CC hybridoma, which produces an antibody useful for ovarian cancer
CC assessment; and (3) inhibiting ovarian cancer in a patient. AAF98573 to
CC AAF98593 represent human kinase marker primers and probes which are used
CC in the exemplification of the present invention
XX
XX Sequence 1835 BP; 309 A; 621 C; 527 G; 378 T; 0 U; 0 Other;
Qy
Qy Query Match 21.1%; Score 232.4; DB 5; Length 1835;
Qy Best Local Similarity 55.0%; Pred. No. 4.8e-32;
Qy Matches 508; Conservative 0; Mismatches 401; Indels 15; Gaps 2;
Qy 1 GGGCCCTTGTCTCTGGCCATGCGCCAGAGGGGTCTCTGGGGCTTGGGCGAGCTGGGGGT 60
Db 212 GGGCCCTTGTCTCTGGCCATGCGCCAGAGGGGTCTCTGGGGCTTGGGCGAGCTGGGGGT 271
Qy 61 GTGGCCAAATTCAGACTCATCTCACTTACGGGTGTGTGCGTCCGGACCCCGTAGGGGC 120
Db 272 GTGGCCAAATTCAGACTCATCTCACTTACGGGTGTGTGCGTCCGGACCCCGTAGG 331
Qy 121 CCGCGTACTGCGGGCGCCCTGAGCCCTCGGCCCGCATCTGGGGGGCTCAACCGCGAG 180
Db 332 GCTCCCTGGGTGTG-----GCCCGCCAGCACCATCAAGTGTGGCAGTGCAGTTC 385
Qy 181 CCGGGCACCTGGCCCTTGGCAAGTGAAGCCTGCACCATGAGGTGGCCACATCTGCGGGGC 240
Db 386 GCGGTGAGTGGCCCTGGCAGTCAAGTCACTATGAAGGGTCCATGTGTGTGTGGTGC 445
Qy 241 TCCCTCATGCGCCCTCTCTGGTCTCTCTGGTCTCTCTGGTCTCTCTGGTCTCTCTGGTCT 300
Db 446 TCTCTGTGTCTGAGCAGTGGGTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 498
Qy 301 TTGGAGCCCGCGCGAGTGGTGTGCTGTCTGTGGGCTGCACTCCACAGAGCGGGCCCTG 360
Db 499 --GCACCAAGGAAGCCATAGAGTCAAGCTTGGGGGCCACACAGTACTCTACTCTCC 556
Qy 361 GACGGCGGCAACACCGGGCAGTGGCCCGCATCTGTGGTGTGGCGCCCAACTACAGCCAAAGT 420
Db 557 GAGGAGCCCAAGGTTCAGCACCTCTGAAGGACATCATCCCGCCACCCAGCTACTCTCCAGAG 616
Qy 421 GAGCTGGGGCCGAGCTGGCCCTGTGCGCTGTGGCTCACCGCCAGCTGGGCCCGCCGCC 480
Db 617 GGCCTCCAGGGGCAATTCGACTCTCTCAACTTCAGCAGACCCATCACTTCTCCCGTAC 676
Qy 481 GTGTGGCTGTCTGCTGCGCCCGCCCTCACACCGCTTGTGTCAGCGACCCCGCTGTGG 540
Db 677 ATCCGGCCCATCTGCTCTGCTGAGCCAGGCTCTCTTCCCAACAGGCTCTCCACTGCACT 736
```


541 GCACCGCGCTGGGAGACGCTCCAGGAGGACAGATCCTCTGCTCTCTCCCTGGGTGCTACAG 600
 737 GTCACTGGCTGGGTCAATGTGGCCCTCTCACTGAGCTCTGAGCCCAAGCCACTGCAG 796
 601 GAAGTGAGCTAAGGCTGTGGGAGGAGGACCTGTCTCAATGTCTTACAGCCAGCCCGGT 660
 797 CAATCGAGGTGCTCTGATCATGTCGAGACGCTGTAACTGCTGTAAACATCGAGCC 856
 661 CCCTTCAACCTCACTCTCCAGATATTCAGGATGCTGTGCTGCTACCCAGAGGGC 720
 857 AAGCTGAGGAGCGCACTTTGTCCAGAGACATGTTGTGTGTGTGTGTGTGTGTGTGTGT 916
 721 CGCAGGACACCTCCAGGCTGACTCTGGGGGGCCCTGTGTGTGTGTGTGTGTGTGTGTGT 780
 917 GGCAGAGCGCTCCAGGCTGACTCTGGGGGGCCCACTCTCTCCCTGTGTGTGTGTGTGT 976
 781 TGGTTCAGGAGGAATCACAGCTTTGGGTTTGGCTGTGTGTGTGTGTGTGTGTGTGTGT 840
 977 TGGTACCTGACGGGCAATGTGAGCTGGGGAGATGCTGTGTGGGGCCCGCAACAGGCTGGT 1036
 841 GTTTTCACTGCTGTGCTTACCTATGAGGACATGATAGGAGGAGGAGGAGGAGGAGGAGG 900
 1037 GTGTACACTGTGGCTCGAGCTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1096
 901 CTGGGGCTGCTCTTTCCACCCAG 924
 1097 CCTCGTGTGTGGTCCCAACCCAG 1120
 RESULT 13
 ID ABZ35336
 XX ABZ35336 standard; cDNA; 3382 BP.
 AC ABZ35336;
 XX 05-FEB-2003 (first entry)
 XX Human gene expression profile polynucleotide SEQ ID NO 447.
 KW Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;
 KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;
 KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;
 KW gene expression; gene; ss.
 XX Homo sapiens.
 OS
 XX WO200274979-A2.
 PN
 XX 26-SEP-2002.
 PD
 XX 20-MAR-2002; 2002WO-US008456.
 PF
 XX 20-MAR-2001; 2001US-0276947P.
 PR
 XX (ORTH) ORTHO CLINICAL DIAGNOSTICS INC.
 XX
 XX Wan J, Wang Y;
 XX WPI; 2002-740862/80.
 XX
 XX New gene expression profile generated from primary, endothelial,
 PT epithelial, and muscle cell types, useful for identifying disease
 PT pathologies involving alterations of gene expression, e.g. cancer.
 XX
 XX Example 3; Page 595-496; 850pp; English.
 XX
 XX The invention relates to a gene expression profile comprising one or more
 CC genes (ABZ3489-ABZ35692) and generated from a cell type. The cell type
 CC is a coronary artery endothelium, umbilical artery or vein endothelium,
 CC aortic endothelium, dermal microvascular endothelium, pulmonary artery
 CC endothelium, myometrium microvascular endothelium, keratinocyte
 CC epithelium, bronchial epithelium, mammary epithelium, prostate
 CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,

CC small airway epithelium, renal epithelium, umbilical artery smooth
 CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,
 CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,
 CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,
 CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,
 CC osteoblasts or prostate stromal cell. The gene expression profile is used
 CC for determining the level of RNA expression for a sample, determining the
 CC phenotype of a cell and distinguishing cell types. The gene or a protein
 CC expression profile is useful in identifying disease pathologies involving
 CC alterations of gene expression. The assessment of expression profiles may
 CC provide meaningful information with respect to tumour type and stage.
 CC treatment methods, and prognosis. The gene or protein expression profile
 CC may also be used for creating microarrays. The microarray is useful for
 CC genetic and physical mapping of genomes, DNA sequencing, genetic or
 CC medical diagnosis, genotyping of organisms, confirming cell or tissue
 CC identifications and in identifying promising antibiotics, antiviral or
 CC antifungal agents
 XX
 SQ Sequence 3382 BP; 602 A; 1069 C; 959 G; 752 T; 0 U; 0 Other;
 Query Match 21.1%; Score 232.4; DB 6; Length 3382;
 Best Local Similarity 55.0%; Pred. No. 4.7e-32;
 Matches 508; Conservative 0; Mismatches 401; Indels 15; Gaps 2;
 Qy 1 GGGCCCTTGTCTGGCCATGCGCCCAAGAGGGGCTCTGGGGCTGGGCGAGCTGGGGGCT 60
 Db 212 GGGCCCTTGTCTGGCCATGCGCCCAAGAGGGGCTCTGGGGCTGGGCGAGCTGGGGGCT 271
 Qy 61 GTGGCCCAATTCGACTCATACTTACGGTGTGGTCCGTCGAGCCCGCTAGGGGC 120
 Db 272 GTGGCCCAATTCGACTCATACTTACGGTGTGGTCCGTCGAGCCCGCTAGGGGCAGAA 331
 Qy 121 CCCCCTGACTCGGGGCGCCCTCGAGCCCTCGGCCCGCATCTGGGGGGCTCAAAAGCGCAG 180
 Db 332 GCTCCCTGGCGTGTG-----GCCCCCAAGCAGCATCACAGTGGCAGCAGTCAGTC 385
 Qy 181 CCGGGCACTGGCTTGGCAAGTGAGCTGCACATGAGAGGTGGCCACATCTGGGGGGC 240
 Db 386 GCGGTCAGTGGGCTGGCAGGTGAGCATCATCTATGAAGGCGTCCATGTGTGTGTGGC 445
 Qy 241 TCCCTCATCGCCCTCTCGGTCTCTCGCTGCTCTCCTGCTGCTGCTGCTGCTGCTGCTGCT 300
 Db 446 TCTCTCTGTGTGAGCAGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 498
 Qy 301 TTGAGCCCGCGCCGAGT 360
 Db 499 --GCACCACAAGGAAGCCTATGAGGTCAAGCTGGGGGCCCAACAGCTAGACTCTCTCTCC 556
 Qy 361 GACGGCGGCACACCCCGCAGTGGCGCCCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
 Db 557 GAGGAGCCCAAGGTTCAGCACCCCTGAAGGACATCATCCCCCAACCCAGCTCTCTCTCTCT 616
 Qy 421 GAGCTGGGCGCGACCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 Db 617 GGCTCCAGGGGACATTTGACATCTCTCACTGAGCAGACCATCTCTCTCTCTCTCTCTCTCT 676
 Qy 481 GTGTGGCTGT 540
 Db 677 ATCGGCCCATCTGCTCTCCCTCGCAGCCAAAGCCTCTTCCCAACAGCCCTCTCTCTCTCT 736
 Qy 541 GCGACCGGCTGGGAGACGCTCCAGGAGGAGATCTCTGTGCTCTCTCTCTCTCTCTCTCTCT 600
 Db 737 GTCACTGGTGGGTCTATGTGGGCCCTCTGAGGAGCTCTCTGAGCCCAAGCCACTCTGAG 796
 Qy 601 GAAGTGGAGCTAAGGCTGTGGGGAGGCCACCTGTCTCAATGTCTCTCTCTCTCTCTCTCTCT 660
 Db 797 CAATTCAGGTGCTCTGTATCATGTCGTGAGACGTGTAACTGCTGTGTGTGTGTGTGTGTGT 856
 Qy 661 CCCTTCAACCTCACTCTCCAGATATTCAGGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGT 720
 Db 857 AAGCTGAGGAGCGCACTTTGTCCAAAGAGACATGTTGTGTGTGTGTGTGTGTGTGTGTGT 916
 Qy 721 CGCAGGAGACCTGCGAGGAGTGTCTGTGGGGGGCCCTGTGTGTGTGTGTGTGTGTGTGTGT 780

Db 917 GGCAGAGCGCTGCGAGGTGATCTTGGGGGCCCACTCTCTGCTTGGAGGTCTC 976
Qy 781 TGGTTCAGGAGGAAATCACAGCTTTGGGTTGGCTGTGGACGGAGAAACCGCCCTGGA 840
Db 977 TGGTACCTGACGGGCATTTGAGCTGGGGAGATGCTGTGGGGCCCGCAACAGGCTGGT 1036
Qy 841 GTTTTCACTGCTGGTACCTATGAGCATGAGTACGGGAGCAGGTGATGGTTTCAGAG 900
Db 1037 GTGTACACTCTGGGCTCAGTATGCTCTCTGGATCCAAAGGAGTACAGAACTCCAG 1096
Qy 901 CCTGGGCTGCTTTTCCACCCAG 924
Db 1097 CCTGCTGGTGCCCAACCCAG 1120

RESULT 14
ABT31936
ID ABT31936 standard; DNA; 1733 BP.
XX AC ABT31936;
XX DT 01-MAY-2003 (first entry)
XX DE Human breast cancer / ovarian cancer related coding sequence #43.
XX KW Human; gene; ds; cytostatic; breast cancer; ovarian cancer.
XX OS Homo sapiens.
XX FN WC2003000012-A2.
XX PD 03-JAN-2003.
XX PF 21-JUN-2002; 2002MO-US019773.
XX PR 21-JUN-2001; 2001US-0300159P.
XX PR 27-JUN-2001; 2001US-0301351P.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PI Veiby OP;
XX DR WPI; 2003-267848/26.
XX DR P-PSDB; ABJ37067.
XX PT Determining the presence of breast cancer in an individual, involves
XX PT using specific polynucleotide markers.
XX PS Disclosure; Page 192-193; 233pp; English.
XX CC The invention comprises a method for assessing whether a patient is
XX CC afflicted with breast cancer or ovarian cancer. The method involves the
XX CC use of specific DNA markers. The method of the invention is useful in the
XX CC detection and treatment of ovarian and breast cancer. DNA sequences
XX CC ABT31894 - ABT31949 encode human breast/ovarian cancer-related proteins
XX SQ Sequence 1733 BP; 311 A; 578 C; 500 G; 344 T; 0 U; 0 Other;

Query Match 20.9%; Score 230.8; DB 7; Length 1733;
Best Local Similarity 54.9%; Pred. No. 9.2e-32;
Matches 507; Conservative 0; Mismatches 402; Indels 15; Gaps 2;

Qy 1 GGGCCCTTGTCTCGGCGCATGCCCCAGAGGGGTCCTGGGGCCTGGGAGCTGGGGGCT 60
Db 81 GGGCCCTTGTCTCGGCGCATGCCCCAGAGGGGTCCTGGGGCCTGGGAGCTGGGGGCT 140
Qy 61 GTGGCCAAATTCGACTCACTTACGGGTTGGTGGCTCCGGACCCCGCTAGGGGC 120
Db 141 GTGGCCATTCCTCTATCTTGGATTACTCCGGTCAGGGACAGAGGCGGAGGGGCAGAA 200
Qy 121 CCCCCGTACTCGGGCGCCCTCAGGCCCTCGGCCCGCATGTCGGGGGCTCAACCGCAG 180

Db 201 GCTCCCTGGGGTGTG-----GCCCCCAAGCAGCATCAAGTGGCAGCAGTGCAATC 254
Qy 181 CCGGCACTTGGGCTTGGCAAGTGAGCTTGCACCAATGGAGGTGGCCACATCTTGGGGGGC 240
Db 255 GCGGTCAGTGGGCGCTGGCAGGTGAGCATCACCTATGAAGGCGTCCATGTGTGTGGTGGC 314
Qy 241 TCCCTCATCCGCCCTCTCTCGCTGCTCGCTGCTCAGTGTCTCAGTGTCTTTCATGCAATGGGAGC 300
Db 315 TCTCTGCTGTGAGCAGTGGGTGCTGTGAGTGTCTCAGTGTCTCAGTGTCTTCCCAGGA----- 367
Qy 301 TTGAGCCCCCGGCGGAGTGGTCTGGTACTCTCTGGGCGTGCACCTCCAGAGCGGGCCCTG 360
Db 368 --GCACCAAGGAAGCCTATGAGGTCAAGCTGGGGGCCACACAGCTAGACTCTCTACTCC 425
Qy 361 GACGGCGGCACACCCGGCAGTGGCCGCTATGCTGTGTGTGGCCCACTACAGCAAGTG 420
Db 426 GAGGACGCCAAGGTGAGCACCCCTGAAGGACATCATCCCCCACCCAGCTACCTCCAGGAG 485
Qy 421 GAGCTGGGCGCGGACCTTGGGCTGCTGCGCTGGCTCAGCCGCGCAGCTGGGGCCCGGCC 480
Db 486 GGCTCCAGGGGAGACATTTGACCTCTCCAACTCAGCAGACCCATCACTTCTCCGGTAC 545
Qy 481 GTGTGGCTGTGTGGCTGGCCGGCGCTCACAACGCTTGTGTGCAAGGACCGCTGTGCG 540
Db 546 ATCGGGCCCATCTGCTCTCCCTGACGCCAAGCCCTCTTCCCAACAGGCGCTCCACTGCACT 605
Qy 541 GCCACCGCTGGGGAGACGTCCAGAGGAGATCTCTGCTCTCTCCCTGGGTCTACAG 600
Db 606 GTCACTGGCTGGGGTCAATGTGGGCCCCCTCAGTGAAGCTCTCTGACGCCCAAGCCACTG 665
Qy 601 GAAGTGAGCTAAGGCTGTGGGCGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGT 660
Db 666 CACTCGAGGTGCTCTGATCAGTCTGTGAGCGTGTAACTGCTGTACAACTACAGCGCC 725
Qy 661 CCCTTCAACCTCACTTCCAGATATGCCAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
Db 726 AAGCCTGAGGAGCGCACTTTGTCCAAAGAGACATGTGTGTGTGTGTGTGTGTGTGTGTGT 785
Qy 721 CGCAGGACACCTGCCAGGTGACTCTGGGGGCGCCCTGGTCTGTGAGGAGCGGCGCC 780
Db 786 GGCAGGACGCTGCCAGGTGACTCTGGGGGCGCCACTCTCTGCTGTGTGTGTGTGTGTGTGT 845
Qy 781 TGGTTCAGGAGGAATCACAGCTTTGGGTTTGGCTGTGAGCGAGAGAAACCGCCCTGGA 840
Db 846 TGTACCTGACGGGCATTTGTGAGCTGGGAGATGCTGTGGGGCGCCGCAACAGGCTGGT 905
Qy 841 GTTTTCACTGCTGTGGCTACCTATGAGCATGATAGGAGCAGGTGATGGTTTCAGAG 900
Db 906 GTGTACACTCTGGCCTCCAGCTATGCTCTCTGGATCCAAAGCAAGGTGACAGAACTCCAG 965
Qy 901 CCTGGGCTGCTTTTCCACCCAG 924
Db 966 CCTGCTGTGTGCCCCCAACCCAG 989

RESULT 15
ABA94396
ID ABA94396 standard; cDNA; 944 BP.
XX AC ABA94396;
XX DT 26-MAR-2002 (first entry)
XX DE Human prostatic-like serine protease encoding cDNA.
XX KW Prostatic-like enzyme; human; prostatic-like serine protease; cytostatic;
XX KW antiatherosclerotic; virucide; osteopathic; antiinflammatory; vasotropic;
XX KW neuroprotective; gene therapy; antisense therapy; ss.
XX OS Homo sapiens.
XX FH Key
XX CDS Location/Qualifiers
1..819

FT /*tag= a
FT /product= "prolactin-like serine protease"
XX

PN WO200119466-A2.

XX 27-DEC-2001.

XX 22-JUN-2001; 2001WO-EP007116.

XX 23-JUN-2000; 2000US-0213474P.

PR 22-MAR-2001; 2001US-0277612P.

XX (FARB) BAYER AG.

XX PA

XX Xiao Y;

XX WPI; 2002-114575/15.

DR P-PSDB; ABB07286.

XX Novel human prolactin-like enzyme polypeptide and polynucleotide which
PT can be regulated for treating metastasis of malignant cells,
PT inflammation, atherosclerosis, neurodegenerative disease and pathogenic
PT infection.

XX Claim 1; Fig 5; 125pp; English.

XX The invention relates to human prolactin-like enzyme polypeptides and
CC polynucleotides. The enzyme can be expressed by standard recombinant
CC methodology. The polypeptide, polynucleotide and modulators are useful
CC for treating diseases like metastasis of malignant cells, tumour
CC angiogenesis, inflammation, chronic obstructive pulmonary disease (COPD),
CC atherosclerosis, neurodegenerative disease and pathogenic infection,
CC particularly viral infection. The prolactin-like enzyme gene provides a
CC therapeutic target of decreasing the enzyme activity, in particular for
CC treating or preventing metastatic cancer. Neurodegenerative diseases
CC include for e.g. prion protein amyloid plaques of Genstmann-Straussler
CC Syndrome, Creutzfeldt-Jakob disease and Scrapie. The agonists and
CC antagonists of the polypeptide may be useful to treat osteoporosis,
CC Paget's disease, degradation of bone implants particularly dental
CC implants. Altered levels of human prolactin-like enzyme activity inhibit
CC both smooth muscle cell proliferation and lipid accumulation and inhibit
CC the progression of restenosis and atherosclerosis. Anti-human prolactin-
CC like serine protease antibodies are useful for immunodetection and
CC diagnosis of micrometastases, autoimmune lesions and renal failure in
CC biopsy specimens, plasma samples and body fluids. The present sequence
CC represents a cDNA encoding a human prolactin-like serine protease

XX Sequence 944 BP; 150 A; 318 C; 318 G; 158 T; 0 U; 0 Other;

Query Match 20.5%; Score 226.4; DB 6; Length 944;
Best Local Similarity 57.7%; Pred. No. 5.6e-31;
Matches 448; Conservative 0; Mismatches 316; Indels 12; Gaps 2;

QY 124 CCGTACTGGGGGCGCCCTGAGCCCTCGGGCCGATCGTGGGGGGCTCAACGGCGAGCGG 183

Db 43 CTGGCCCTGGGGCGAGCCCGCCGATGTCAGTGGGATGCTGGGGGGCGGGATGCGCGGAC 102

QY 184 GGCACCTGGGCTTGGCAAGTGGAGTGCACCATGGAGGTGGCCACATCTCGGGGGGTCC 243

Db 103 GGAGAGTGACCTGGGAGGCGAGGATCAGCATCGTGGGGGACACATGCTGGGGGGGTGG 162

QY 244 CTATCGCCCGCCCTCTGGGTCTCTCGGCTGCTCACTGTTTTCATGACGAATGGAGCGTTG 303

Db 163 CTCATGCGCCCGCCCGCTGCTGACAGCGGCGCACTGCTTCCCGAGGA-----GG 213

QY 304 GAGCCGGGGCGGAGTGGTGGTACTGCTGGGGGTGCACTCCAGAGAGCGGCCCTCGAC 363

Db 214 GCACTGGCAGCTGAGTACCGGCTGGGGCGGTGGGTCTGGGCTCCACCTCGCCCC 273

QY 364 GGGCGGACACCCCGGAGTGGCGCCCATCTGGTGGCGGCGCAACTACAGCAAGTGGAG 423

Db 274 CGCAGCTCTCGGTGCGCGTGGAGGGGTGCTGCTGCCCGCGGACTACTCGAGGACGGG 333

QY 424 CTGGGCGCGGACCTGGCCCTGCTGGCCCTGCGCTCACCGCCAGCCTGGGGCCCCCGCTG 483

Db 334 GCCCGCGGAGACCTGGCACTGCTGAGCTGCGTGGCGGGTGGCCCTGAGCGCTCGCGTC 393

QY 484 TGGCTCTGCTGGCTGGCCCGCGGCTTCAACCGCTTGGTGGCAGGGACCGCTGCTGGGGCC 543

Db 394 CAACCCGCTGCTGCTGCGGCTGCGCGGCGCCCGCGCGCGCGGACACCAATGCGGGGTC 453

QY 544 ACCGGCTGGGAGAGCTCCAGGAGCGAGATCTCTGCGCTCTCCCTGGGTGCTACAGAA 603

Db 454 ACCGGCTGGGAGAGCTCCAGGAGCGAGATCTCTGCGCTCTCCCTGGGTGCTACAGAA 513

QY 604 GTGGAGCTAAGGCTGCTGGGCGAGGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663

Db 514 GTAAAGGCTGCGGCTGCTGGAATCTCGGCGGAGCTGCGGCGGAGCTTACCACTGGGCGGAG 573

QY 664 TTCAACCTCACTCTCCAGATAT---TGGCAGGAGTGTGTGCTGGGTACCCAGAGGCG 720

Db 574 GTGGCCCGAGGCTGAGCGCATTTGCTGCTGGGAGTCTGTGTGGCGGCTTACCCCGAGGCG 633

QY 721 CGCAGGAGACACCTGCGGAGGCTGCTGCTGGGGGGCGCCCTGGGTCTGTGAGGAGCGGCGCG 780

Db 634 CACAGAGGAGCGCTGCGGAGGCTGCTGCGGAGGCTCTGACCTGCTGCTGCTGCTGCTGCT 693

QY 781 TGGTTCCAGGCGAGGAATCACAGCTTTGGGTTTGGCTGTGGACGGAGAAACCGCCCTGGA 840

Db 694 TGGGTCTGCTGGGCGTGGTGGAGTGGGCGGAGGCTTGTGCGCTGCGGCGGAGGCTCCAGGG 753

QY 841 GTTTTCACTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 896

Db 754 GTCTACACAGTGTGGCGACATATAGCCCTGGATTTCAGGCTCGGCTCGGCTCAAGATTAC 809

Search completed: February 25, 2004, 09:18:34
Job time : 482 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 08:20:46 , Search time 2783 Seconds

(without alignments)
11824.694 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102

Sequence: 1 gggccctgtctcgggcat.....ggggtttctgatgggctctc 1102

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_estci.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rtd.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gsl.*

29: gb_gsl2.*

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
C 1	533.8	48.4	537	10	AW450407
C 2	442.8	40.2	456	9	AI190509
C 3	214.2	19.4	1071	9	AL551470
C 4	202	18.3	433	10	BB850564

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
AW450407/c
LOCUS
DEFINITION
UI-H-BI3-akn-g-11-0-UI-s1 NCI CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2735037 3', mRNA sequence.
ACCESSION
AW450407
VERSION
AW450407.1 GI:6991183
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 537)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
Oligo-dr track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
information can be found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnl.gov/bbrp/image/image.html
Seq primer: M3 Forward
POLYA=No.

FEATURES

Location/Qualifiers

BB850607 BB850607
AK078696 Mus muscu
BM923713 AGENCOURT
BB850795 BB850795
AK010640 Mus muscu
BI763558 603050292
BY710051 BY710051
BB849543 BB849543
BX436299 BX436299
BB593614 BB593614
AZ252377 RPCI-23-4
AK080281 Mus muscu
BC034294 Homo sapi
AL555870 AL555870
BY234907 BY234907
AK06271 Mus muscu
BG762809 602734736
BM917234 AGENCOURT
AL578261 AL578261
CD389322 AGENCOURT
AK004939 Mus muscu
BY235958 BY235958
CB851297 UI-CF-EC1
BY236911 BY236911
BI259237 602970095
BX417595 BX417595
CA489381 AGENCOURT
BQ690230 AGENCOURT
BE590187 197182 BA
CD050808 AGENCOURT
CE450243 tigr-gss-
BY706352 BY706352
BI554641 603235988
BT768651 603056994
BX644781 DKFZp781G
CB794592 K-EST0075
CB995955 AGENCOURT
BY236668 BY236668
CF900000 A0310C05-
CB054579 NISC sm05
AK014645 Mus muscu

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Nomura,M., Matsuyama,T., Nakamura,M., Niishi,K., Nomura,K., Nomasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sugabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,Y., Tomaru,A., Toyota,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL COMMENT Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
waga,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,F., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.
Location/Qualifiers
1. .433
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F930108A18"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"

FEATURES
source

ORIGIN

Query Match 18.3%; Score 202; DB 10; Length 433;
Best Local Similarity 81.0%; Pred. No. 5.9e-46;
Matches 235; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 84 ACTTTTACGGGTGTGTGGCTCCGACCCGCTAGGGGCCCCCGCTACTGTGGGGCGCCCTGA 143
Db 141 ACTCAGTGTGTGACTCTCTACCCAAAGAGAAATTGAGACCTAGACTGCGGCGCCCTGA 200

QY 144 GCCCTCGGCCCGCATCGTGGGGGCTCAAAAGCGACGCGGGCACTCTGGCCTTTGGCAAGT 203
Db 201 GCCCTCTTCCCGCATTTGTGGGGGCTCGGAACGCTCATCTGGCACTTGGCCATGGCAGGT 260

QY 204 GAGCTGCAACATGAGGTGGCCACACTGCGGGGGCTCCCTCATGCCCCCTCTCTGGGT 263
Db 261 GAGCCTGTGATCAAGTGGGGGCCACATCTGGGGGGGTCCCTCATGCCCTCTCTGGGT 320

QY 264 CCTCTCCGCTGCTCACTGTTTCATGACGAATGGGACGTTGGAGCCCGCGCGAGTGTC 323
Db 321 CCTCTCCGCTGCTCAATGTTTCGTGTCGAATGGAACTTGGAGCCCGCGAGAAATTGC 380

324 GGTACTGCTGGCGGTGCATCCACAGACGCGCCCTGGAGCGCGCAC 373
 |||
 381 AGTTCTGCTGGCGGTGCATCCACAGACGCGCCCTGGAGCGCGCAC 430
 |||

RESULT 5
 BB850607 433 bp mRNA linear EST 26-NOV-2001
 LOCUS
 DEFINITION BB850607 RIKEN full-length enriched, adult inner ear Mus musculus
 cDNA clone F930108E12 5', mRNA sequence.
 ACCESSION BB850607
 VERSION
 KEYWORDS BB850607.1 GI:17092061
 EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 433)
 AUTHORS Akimura,T., Arawaka,T., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
 Hayatsu,N., Hiromoto,K., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
 Ishii,Y., Ito,M., Kawai,J., Nishi,K., Nomura,K., Numasaki,R.,
 Matsuyama,T., Nakamura,M., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
 Okazaki,Y., Okido,T., Saito,R., Shindagawa,A., Shiraki,T.,
 Sasaki,D., Sato,K., Shibata,K., Shinaga,A., Shiraki,T.,
 Sogabe,Y., Tomaru,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
 Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)
 TITLE Unpublished (2001)
 JOURNAL Contact: Yoshihide Hayashizaki
 COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
 Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
 and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
 Sugahara,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.
 FEATURES Location/Qualifiers
 source 1. 433
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="F930108E12"
 /tissue_type="inner ear"
 /dev_stage="adult"
 /clone_lib="RIKEN full-length enriched, adult inner ear"

Query Match 18.1%; Score 199.2; DB 10; Length 433;
 Best Local Similarity 86.9%; Pred No. 1.9e-25;


```

Qy 541 GCCACCGCTGGGAGAGCTCCAGGAGCGACATCTCTGCTCTCCCTGGTGGTCTACAG 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 620 GTCACTGGCTGGGTCTATGTGGCCCCCTCATGAGCCTCTCTGACGCCACCACTGAG 679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 601 GAAGTGAGCTAAGGCTGCTGGGCGAGGCCACCTGTCTCAATCTCTCAAGCCCGCGT 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 680 CAACCTCGAGGTGCCTCTGATCAGTCTGTGAGACGTGTAACTGCCTGTATCAACATCGAGCC 739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 661 CCTTTCACCTCACTCTCCAGATATGCGAGGATGCTGTGTGCTGTGCTACCCAGAGGGC 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 740 AAGCTGAGGAGCGGCACATTTTTCACAGAGACATGGTGTGCTGTGCTATGTGAGGGG 799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 721 CGCAGGACACCTGCCAGGGTGACTCTGGGGGGCCCC 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 800 GGCAAGGAGCGCTGCCAGGGTGACTCTGGGGGGCCCC 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 8
BB850795 433 bp mRNA linear EST 26-NOV-2001
LOCUS BB850795 RIKEN full-length enriched, adult inner ear Mus musculus
DEFINITION BB850795.1 GI:17092249
ACCESSION BB850795
VERSION BB850795.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)

```

```

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1. (bases 1 to 433)
Akimura,T., Arawaka,T., Hiramoto,K., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)

```

```

JOURNAL
COMMENT Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
Location/Qualifiers
1. .433

```

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F930109F10"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN

Query Match 17.4%; Score 191.2; DB 10; Length 433;
Best Local Similarity 86.7%; Pred. No. 4.9e-24;
Matches 222; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

Qy 123 CCCGTACTCTGGGGCGCCTCGAGCCCTCGGCCCGCATCTGGGGGGGCTCAAAACGGCGCAGCC 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 CCTAGACTGGGGCGCCTCGAGCCCTCTCCGCAATTGTGGGGGGCTCGGACGCTCATCC 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 183 GGGCACCTGCGCTTGGCAAGTGAGCTGCAACATGAGAGGTGGCCACATCTGCGGGGGGCTC 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 TGGCACCTTGGCCATGGCAGGTGAGCCTGTCATCAAGGTGGGGGCCACATCTCGGGGGGGCTC 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 243 CCTCATCGCCCCCTCTGGGGTCTCTCGGCTGCTCACTGTTTCATGACGAATGGACGTT 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 CCTCATCGCCCCCTCTCTGGGT-CTCTCCGCTGCTCACTGTTTCGAGCAATGGAACCTT 356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 303 GGAGCCCGCGCGAGTGGTGGTCTGCTGCGGTGCTGCGGGGCTGCACTCCCGAGACGGGGCCCTGGA 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 GGAGCCCGCGGAGAAATGTTCAGTTCTGCTGGGGGTGCACTCCCGAGACGGGGCCCTGGA 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 363 CGGCGCGCACACCGCGC 378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 AGGAGCGCACATGGGC 432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AKO10640 1629 bp mRNA linear HTC 20-SEP-2003
LOCUS AKO10640 Mus musculus ES cells cDNA, RIKEN full-length enriched library,
DEFINITION clone:2410039E18 product:PROTASIN, full insert sequence.
ACCESSION AKO10640.1 GI:12846228
VERSION AKO10640.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1. Carninci,P. and Hayashizaki,Y.
REFERENCE
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
TITLE
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
TITLE
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913

```


ACCESSION BI763558
VERSION BI763558.1 GI:15755136
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 861)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rcapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L14M11475 row: h column: 23
High quality sequence stop: 861.
Location/Qualifiers
1..861
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5190214"
/lab_host="DH10B"
/clone_lib="NIH_MGC_116"
/note="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dr primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen) Research Genetics tracking code
023. Note: this is a NIH_MGC Library."

Query Match 15.9%; Score 175; DB 12; Length 861;
Best Local Similarity 52.3%; Pred. No. 4.6e-21;
Matches 444; Conservative 0; Mismatches 390; Indels 15; Gaps 2;
QY 1 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGGTCTTGGGGCCCTGGGCAGCTGGGGCT 60
DB 2 GGGGCTTGTCTGGGCGCATGGCCAGAGGGGGTCTTGGGGCCCTGGGCAGCTGGGGCT 61
QY 61 GTGGCCAACTCTGACTCATCTCACTTTACGGGTTGGTGGCGTCCGGACCCGCTAGGGGC 120
DB 62 GTGGCCAACTCTGACTCATCTCACTTTACGGGTTGGTGGCGTCCGGACCCGCTAGGGGC 121
QY 121 CCGCGTACTGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAAAACGCGAG 180
DB 122 GCTCCCTGGGTGTG-----GCCCGCCCAAGCAGCATCACAGGTGGCAGGTGCACTC 175
QY 181 CCGGCGACCTTGGCAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 240
DB 176 GCGGCTGAGTGGCCCTGGCAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 235
QY 241 TCCCTCATCGCCCTCTCTGGGTCTCTCGGCTGCTCACTGTTTCATGCAAGTGGGACG 300
DB 236 TCTCTGCTGTGAGCAGTGGGTGTGTGAGTGTCTCACTGCTTCCCGACGCA----- 288
QY 301 TTGGAGCCCGGGCGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
DB 289 --GCACCAAGAGAGCCTATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 346
QY 361 GACGGCGGCGACACCGGCGGCGAGTGGCGGCGATCGTGGTGGCGGCGCACTAGCGCAAGTG 420

DB 347 GAGGAGCCCAAGGTTCAGCACCTTGAAGGACATATCCCCACCCAGCTACCTCCAGGAG 406
QY 421 GAGCTGGGGCGGACCTGGCCCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 480
DB 407 GGTCTCCAGGGCGGACATTCGACCTCTCCAACTCAGCAGAGCCCATCCTTCTCCCGTAC 466
QY 481 GTGTGGGCTGTCTGCTGGCGCCCGGCGCTCACCGCTTGTGGTGGCAGCGCACCGCTGGTGG 540
DB 467 ATCCGGCCCATCTGCTCTCCCTCGCAGCCCAAGCGCTCTTCCCAAGCGCTCCCACTGCAC 526
QY 541 GCACCGGCTGGGGAGAGCTCCAGGAGGAGATCTCTGCTCTCCCTGGGTGGTGGTGGTGG 600
DB 527 GTCACTGGCTGGGGTCAATGTGGCCCTCAGTGGAGCCTCTCTGAGCGCCCAAGCACTGCAG 586
QY 601 GAAGTGAGCTAAGGCTGTGGGGGAGGCGCACCTGTCAATGTCTCTACAGCAGCGCCGGT 660
DB 587 CAATCGAGCTGCTCTGATCAGTCTGAGAGCTGTAACTGCTGCTGCTGCTGCTGCTGCTG 646
QY 661 CCCTTCAACCTCACTCTCCAGATATTGCCAGGGATGCTGTGTGCTGGCTACCCAGAGGCG 720
DB 647 AAGCTGAGGAGCGGCACTTTGTCTCAAGAGGACATGCTGTGTGCTGTGTGTGTGTGTGT 706
QY 721 CGAGGAGACCTCCAGGCTGACTCTGGGGGCGCCCTGCTGTGTGAGGAGGCGGCGCGC 780
DB 707 GGCAAGGAGCGCTGCGGAGGTGACTCTGGGGGCGGCACTCTCTTGGCTGTGGAGGGGTC 766
QY 781 TGGTTCCAGGCGGCAATCACAGCTTTGGGTTTGGCTGTGGAGGAGAAACCGCCCTGGA 840
DB 767 TCTGTTACTAGCGGGCTTGTGAGCTGGGAGACATGCTGTGGGGCCCGCAGGCGCTGGT 826
QY 841 GTTTCCTACT 849
DB 827 GTTCTACT 835

RESULT 11
BY710051 1010 bp mRNA linear EST 16-DEC-2002
LOCUS BY710051.1 RIKEN full-length enriched, ES cells Mus musculus cDNA
DEFINITION clone 2410039E18 5', mRNA sequence.
ACCESSION BY710051
VERSION BY710051.1 GI:27121265
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1010)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Bacalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V.,
Ciothia, C.F., Forrest, A., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forni, L.E., Gough, J., Godzik, J., Grimmond, S.,
Gariboldi, M., Gissi, C., Hasegawa, Y., Hasegawa, Y., Hasegawa, Y.,
Guatinich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kutalik, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, J., Ring, B.Z., Ringwald, M.,
Sundelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Walmsley, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kawaguchi, A., Sakai, K., Sasaki, D., Shibata, K.,

Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 PUBMED
 12466851

CONTACT: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Adachi, S., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
 1. 1010
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="2410039E18"
 /cell_type="ES cells"
 /lab_host="SOLR"
 /clone_lib="RIKEN full-length enriched, ES cells"
 /note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',
 GAGAGAGAGATTCGAGTTAAATAATTAATCCCCCCCCCC 3']".

prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATTCGAGTTAAATAATTAATCCCCCCCCCC 3']".

ORIGIN

Query Match 15.9%; Score 174.8; DB 13; Length 1010;
 Best Local Similarity 53.4%; Pred. No. 5.3e-21;
 Matches 469; Conservative 0; Mismatches 387; Indels 22; Gaps 4;
 1 GGGCCCTGTCTGGGCCATGCCAGAGGGGGTCTTGGGGCTGGGCAGCTGGGGCT 60

Db 97 GGGCCCTGTCTGTAGCCATGCCCCCTAAGGGTGGGCTTGGGAGCTTGGGAGCT 156
 QY 61 GTGGCCAAATTCGACTCATCTACATTTACGGTTGGTGGCTCCGAGCCC---GCTAGG 117
 Db 157 GTGACCAATTCCTCTC-----CTTCTCGAATGCTCCAGTCGGGAATCCGAGCTGAC 207
 QY 118 GGGCCCCCGTACTGGGGGGCCCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAACCGG 177
 Db 208 GGGACTGAAGCCCTCTGTGTGTCCTGCTCATCCAGCCAGCATACCGGTGGTGGCAGTGA 267
 QY 178 CAGCCGGGCACTGGCTTGGCAAGTGAAGTGCACCATGAGAGGTGGCCACATCTCTCGGG 237
 Db 268 AAGCCCGGTCACTGGCTTGGCAAGTGCACCATGAGAGGTGGCCACATCTCTCTGGC 327
 QY 238 GGCTCCCTCATCGCCCCCTCTGGGTCTCTCGCTGCTCACTGTTTCATGAGATGGG 297
 Db 328 GGGTGGCTCGTGTCAAAATAAATGGGTGGTGTCTGTCTCACTGCTTCCCCAGAGA---- 383
 QY 298 AGCTTGGAGCCCGGCCCGGAGTGGTCCGTACTGCTGGGGCTGCACCTCCAGGACGGGCC 357
 Db 384 -----ACACAGCAGGGAAGCGTATGAGGTGAAGCTGGGGGCCACAGCTAGACTCTTAC 438
 QY 358 CTGGACGGCGGCGACACCCCGGAGTGGCGGCCCATGCTGGTGGCGGCCCAACTACAGCCAA 417
 Db 439 AGCAATGACACTGTGTGTCACACAGTGGTGCAGATCATCACCACCTCAAGTACCCAGAA 498
 QY 418 GTGGAGCTGGGGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 477
 Db 499 GAGGGTCTCCAGGGGACATCGGCTCATCCGCTCAGCAGTCTCTGCTGCTGCTGCTGCTG 558
 QY 478 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537
 Db 559 TAGATCAGACCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 618
 QY 538 TGGGCCACCGCTGGGGAGACGCTCCAGAGGAGCAGATCTCTGCTGCTGCTGCTGCTGCTGCT 597
 Db 619 ACTGTACCGGATGGGGTCAATGTGGTCTCTCAGTAGAGCTCCAGACCCCTTAGGCTCTG 678
 QY 598 CAGGAAGTGGAGTAAAGTGTCTGGCGGAGGCCACCTGTCAATGTCTCTACAGCCAGCCC 657
 Db 679 CAGCAGTGGAGTACACTCATCAGCCGGGAACCTGTAGTGTCTGTACACATTAAT 738
 QY 658 GTTCCCTCAACCTCACTCTCCAGATATTGCCAGGAGATGCTGTGCTGGCTAGCCAGAG 717
 Db 739 GCGGTGCTGAAGAACCGCACACTATCCAGCAGGACATGCTGTGTGCTGTGCTGTGTAAG 798
 QY 718 GGGCGCAGGAGACCTGCCAGGGTGAATCTGGGGGGCCCCCTGGTGTGTGAGGAGGGCC 777
 Db 799 GGAGGCAAAAGATGCTGCCAGGGTGAATCTGGGGGGCCCCACATCTCTGTACATGGAGGG 858
 QY 778 CGTGTGCTCCAGCAGGAATCACAGCTTTGGTGTGGCTGTGGAGGAGAAACCGCCCT 837
 Db 859 ATCTGTACTTGGCAGGCATTGTGAGTTGGGGTGGATGTGTGTG-TGCCCCAACAGCCT 917
 QY 838 GGAGTTTTCATCTGCTGTGGCTACCTATGAGGCATGGAT 875
 Db 918 GGAGTATACACTGTGACTTCTACTATGCTGCTCCCGAT 955

RESULT 12

BB849543
 LOCUS
 DEFINITION
 cDNA clone F930101E02 5', mRNA sequence.
 BB849543
 BB849543.1 GI:17090997
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 429)

AUTHORS

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

JOURNAL

Unpublished (2001)

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

e mouse tissues.

FEATURES
source

Location/Qualifiers
1..429
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F93010IE02"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN

Query Match 15.6%; Score 171.8; DB 10; Length 429;
Best Local Similarity 84.4%; Pred. No. 1.4e-20;
Matches 205; Conservative 0; Mismatches 37; Indels 1; Gaps 1;
QY 123 CCGTACTCGGGCGCCCTCGAGCCCTCGGCCCGCATCGTGGGGGCTCAACGCGCAGCC 182
Db 178 CTAGACTCGGGCGCCCTCGAGACCTCTCTCCGCATGTGTGGGGGCTCGAGCGCTATCC 237
QY 183 GGGCACCTGGCCCTTGG-CAAGTGAGCGCTGCACATGGAGGTGGCCACATCTCGGGGGCT 241
Db 238 TGGCATTGGCCATGGCCAGGTGAGCTTCATAAAGGTGGGGGCCAAATCTCGGGGGCT 297
QY 242 CCGTATCGCCCTCTCTGGGTCTCTCGCTGCTCAGTGTTCATGAGCAATGGACGT 301
Db 298 CCGTATCGACCTTCTCTGGGTCTCTCGCTGCTCAGTGTTCGTTGAGCAATGGAACCT 357
QY 302 TGGAGCCCGCGCGAGTGTGTGGTACTGCTGGGGTGCACCTCCAGGACGCGGCCCTGG 361
Db 358 TGAGCCCGCGGACGAATTGTGAGTGTGCTGGGGTGCACCTCCAGGACGCGGCCCTGGA 417

QY 362 ACQ 364

Db 418 AGG 420

RESULT 13

LOCUS

DEFINITION BX436299 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP001YG17
3-PRIME, mRNA sequence.

ACCESSION

VERSION BX436299

KEYWORDS

EST.

SOURCE

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 930)

Full-length cDNA libraries and normalization

Unpublished (2001)

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7995.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSOCAP001AD09FM1&cluster=7995.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CSOCAP001AD09FM1.

FEATURES

Location/Qualifiers

1..930

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSOCAP001YG17"

/tissue_type="THYMUS"

/clone_lib="Homo sapiens THYMUS"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

ORIGIN

Query Match 15.6%; Score 171.4; DB 13; Length 930;
Best Local Similarity 99.4%; Pred. No. 2.1e-20;
Matches 172; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 773 GGGCCGCTGGTCCAGGCAGGAATCACACGCTTTGGCTGTGCGAGGAAACC 832
Db 68 GGGCCGCTGGTCCAGGCAGGAATCACACGCTTTGGCTGTGCGAGGAAACC 127
QY 833 GCCCTGGAGTTTTCACCTGCTGCTGCTACCTATGAGGATGATACGGGAGCAGGTGATGG 892
Db 128 GCCCTGGAGTTTTCACCTGCTGCTGCTACCTATGAGGATGATACGGGAGCAGGTGATGG 187
QY 893 GTTCAGAGCCCTGGGCTGCTGCTTCCACCCAGCCCGAGAACCCAGTCAGAT 945
Db 188 GTTCAGAGCCCTGGGCTGCTGCTTCCACCCAGCCCGAGAACCCAGTCAGAT 240

RESULT 14

LOCUS

DEFINITION BB593614 RIKEN full-length enriched, 4 days neonate male adipose
Mus musculus cDNA clone B430104M11 5', mRNA sequence.

ACCESSION

VERSION BB593614

EST.

KEYWORDS

selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies). "

ORIGIN

Query Match	15.3%	Score 168.6;	DB 28;	Length 430;
Best Local Similarity	88.4%	Pred. No. 5.1e-20;		
Matches 183;	Conservative 0;	Mismatches 24;	Indels 0;	Gaps 0;

QY	737	AGGGTGACTCTGGGGGCCCCCTGGTCTGTGAGGAGCGCGCGCTGGTTCCAGGCAGGAA	796
Db	40	AGGGTGACTCTGGGGGACCGCTGGTCTGTGAGGAGCGGCGCGATGTTCTTGGCCGAA	99
QY	797	TCACCAAGCTTTGGTTTGGCTGTGGAGCGAGAAACCGCCCTGGAGTTCACCTGTGTGG	856
Db	100	TCACCAAGCTTTGGCTTTGGCTGTGGAGCGGAGAACCGCCCTGGGGTCTTCACTGCAGTGG	159
QY	857	CTACCTATGAGGCATGGATACGGGAGCAGGTGATGGGTTTCAGAGCCTGGGCTTTC	916
Db	160	CTCCCTATGAGTCGTGGATTTCGGGACACGTGATGGGTTTCAGAACCTGGGCTTTC	219
QY	917	CCACCCAGCCCCAGAGACCCAGTCAG	943
Db	220	CCAGCCAGCTCCAGAGCCACAGTCAG	246

Search completed: February 25, 2004, 11:17:42
Job time : 2791 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 08:23:39 ; Search time 112 Seconds
(without alignments)
5460.320 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102

Sequence: 1 999ccctgtctctggccat.....gggggtctgtatggggcctcc 1102

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	224	20.3	1613	4	US-09-387-375-1
2	216.8	19.7	1130	4	US-09-387-375-8
3	183.2	16.6	1142	4	US-09-386-642-8
4	183.2	16.6	1169	4	US-09-386-642-7
5	165.2	15.0	1110	4	US-09-386-653A-1
6	164.2	14.9	980	4	US-09-023-942A-30
7	162	14.7	1212	4	US-09-620-312D-431
8	161.6	14.7	1100	4	US-09-023-942A-5
9	161.6	14.7	1100	4	US-09-907-794A-256
10	161.6	14.7	1100	4	US-09-905-125A-256
11	161.6	14.7	1100	4	US-09-902-775A-256
12	160	14.5	1081	3	US-09-008-271A-15
13	157.6	14.3	959	4	US-09-023-942A-25
14	154.8	14.0	1094	4	US-09-023-942A-3
15	153.8	14.0	1130	4	US-09-386-653A-8
16	145.8	13.2	1103	4	US-09-386-642-59
17	141.6	12.8	3147	2	US-09-027-337-1
18	141.6	12.8	3147	4	US-09-644-600-1
19	141.6	12.8	3147	4	US-09-644-600-18
20	141.6	12.8	3147	4	US-09-654-600A-1
21	141.6	12.8	3147	4	US-09-654-600A-18
22	140.6	12.8	2152	4	US-09-023-655-157
23	139.8	12.7	736	4	US-09-027-337-1
24	139.8	12.7	1081	2	US-09-016-366A-22
25	139.8	12.7	1081	2	US-08-978-404B-17
26	139.8	12.7	1137	2	US-09-016-366A-18
27	139.8	12.7	1137	2	US-08-978-404B-13

28	138.2	12.5	1128	2	US-09-016-366A-20
29	138.2	12.5	1128	2	US-08-978-404B-15
30	134.6	12.2	771	3	US-03-079-970A-4
31	134.2	12.2	735	3	US-09-079-970A-1
32	133.4	12.1	1154	2	US-09-016-366A-16
33	133.4	12.1	1154	2	US-08-978-404B-11
34	132.4	12.0	1378	4	US-09-907-794A-262
35	132.4	12.0	1378	4	US-09-905-125A-262
36	132.4	12.0	1378	4	US-03-902-775A-262
37	132.4	12.0	1430	4	US-09-386-629-1
38	130.4	11.8	2900	2	US-09-027-337-9
39	130.4	11.8	2900	4	US-09-644-600-9
40	130.4	11.8	2900	4	US-09-654-600A-9
41	126.8	11.5	1166	4	US-09-386-629-2
42	124.4	11.3	1165	4	US-09-023-942A-28
43	123.6	11.2	1553	4	US-03-280-116-10
44	116.6	10.6	1108	2	US-09-016-366A-14
45	116.6	10.6	1108	2	US-08-978-404B-20

ALIGNMENTS

RESULT 1
US-09-387-375-1
; Sequence 1, Application US/09387375
; Patent No. 6485957
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Oi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/09/387,375
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1613
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-387-375-1

Query Match		20.3%	Score 224;	DB 4;	Length 1613;
Best Local Similarity		57.9%	Pred. No. 1.1e-41;		
Matches 442;		Conservative	0;	Mismatches 310;	Indels 12; Gaps 2;
Qy	129	CTCGGGCGGCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAAAACGCGAGCGGGGCAC	188		
Db	149	CTCGGGCGAGCCCGCATGTCAGTCGATCGTTGGGGCGGGATGGCGGGACGGAGA	208		
Qy	189	CTGGCCCTTGGCAGTGGCTGACCATGGAGTGGCCACATCTGGGGGGCTCCCTCAT	248		
Db	209	GTGGCCGTGGCAGCGGAGCATTCAGCATCTGGGGGCACGTTGGGGGGTGGTCTCAT	268		
Qy	249	CGCCCTCTCTGGGGTCTCTCCGCTGCTCTCTTTCATGACGAAAGGGACGTTGGAGCC	308		
Db	269	CGCCCCCAGTGGGTGCTGACAGCGGGCAGCTGCTTCCCAGAA-----GGGCACT	319		
Qy	309	CGGGCGGAGTGGTTCGGTACTGCTGGGGCGTGCACTCCAGGACGGGCCCTTGACGCGCG	368		
Db	320	GCAGCTGAGTACCGGTGCGCTGGGGCGGTGCGCTGGGGTCCACCTGCGCCCGGCAC	379		
Qy	369	GCACACCCGGGAGTGGCCCGCCATCGTGTGGGGCGGCACTACAGCAAGTGGAGTGGG	428		
Db	380	GCTCTCGGTGCGGTGCGAGCGGGTGTCTGCTGCCCGGACTACTCCAGSACGGGGCCCG	439		
Qy	429	CGCCGACCTGGCCCTGCTGCGCTGGCCCTCACCGCCAGCTGGGGCCCGCGCTGTGGCC	488		
Db	440	CGGCGACTGGGACCTGCTCAGCTGGGTGCGCGGGTGGCCCTTGGAGCGCTCGGTGCAAC	499		
Qy	489	TGCTGCTGCGCGCGGCGCTCACACGCTTGTGCAAGGCAACGCGCTGCTGGGCGCACCGG	548		

```

? LENGTH: 1142
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
? OTHER INFORMATION: With homo sapien serine protease catalytic domain
US-09-386-642-8

Query Match          16.6%; Score 183.2; DB 4; Length 1142;
Best Local Similarity 53.5%; Pred. No. 1.6e-32;
Matches 411; Conservative 0; Mismatches 348; Indels 9; Gaps 1

Qy 157 ATCGTGGGGGCTCAAAAGCGCGACCGCGGACCTGTGGCCTTTGGCAAGTAGAGCTGCACCAT 216
      |||||
Db 139 ATCGTGGGGGCTATGTCTAGAGCGCGGTGAGTGGCCCTGGCAGGTGAGCATCACCTAT 198
      |||||

Qy 217 GAGGTGGGCACATCTCGGGGGGTCCTCATCGCCCCCTCTCGGTCTCTCGCTGCT 276
      |||||
Db 199 GAAGGCGTCATGTGTGGTGGCTCTCTCGTGTCTGAGCAGTGGTGTCTGTGACTGCT 258
      |||||

Qy 277 CACTGTGTTTCATGACGAATGGGACGTTGGAGCCCGCGCGAGTGGTTCGTACTGCTGGGC 336
      |||||

```


Db 530 TGCTACCAACCCAAACCAATCAAGAAATGACATGCTGTGCGCGGCTTCGAGAGGGCAA 589
QY 723 CAGGACACCTGCCAGGTGACTCTGGGGGCCCTGGTCTGTGAGGAAGGGCGCGCTG 782
Db 590 GAAGATGCTCGAAGGGGCACTCGGGCGGCCCTGGTGTGCTCGTGGGTGAGTCGTG 649
QY 783 GTTCAGGAGGAATACACAGCTTTGGGTTTGGCTGTGAGCGAGAAACCGCCCTGGAGT 842
Db 650 GCTGAGGGCGGGGTGATCAGCTGGGGTGAGGGCTGTGCGGCCAGAACCGCCAGGTGT 709
QY 843 TTCTACTGCTGTGGCTTACCTATGAGGCATGGATAC 877
Db 710 CTACATCGGTGTCACCGCCCAACCAACTGGATCC 744

RESULT 7

US-09-620-312D-431
; Sequence 431, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; PRIOR FILING DATE: 2000-07-19
; PRIOR FILING DATE: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt FL_genes Version 1.0
; SEQ ID NO 431
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (135)..(1007)
US-09-620-312D-431

Query Match 14.7%; Score 162; DB 4; Length 1212;
Best Local Similarity 53.6%; Pred. No. 9.4e-28;
Matches 414; Conservative 0; Mismatches 340; Indels 18; Gaps 3;
QY 112 GCTAGGGGCCCCCGTACTGCGGGCGGCTGAGCCCTCGCGCATGCTGGGGGGGTCA 171
Db 192 GCCAAGGAGCAACAGCCTGTGGTGGCGCCCGGAGTGTGAACCGAATGTGTGGCGGGCAG 251
QY 172 AACGCGCAGCGGGGACCTCGCCCTTGCAAGTAGTGGCTGACCATGAGGTGGCCATC 231
Db 252 GACACGAGAGGGCGAGTGGCCCTGGCAAGTTCAGCATCAGCGCAACGGAAGCCATTC 311
QY 232 TGGGGGGCTCCCTCATCGCCCCCTCTCGGCTGTCTCACTGTTT---CATG 288
Db 312 TGGGGGGGCGACCTCATCGCGGAGCAGTGGGTCTGTGACGCTGGCACTCTCCGCAAC 371

QY 289 ACGAATGGAGCCTTGAGAGCCCGCGCGAGTGGTGGTACTCTGTGGCGGTGCACTCCCAAG 348
Db 372 ACTCTGAGAGCTCCCTGTGTACAGGTCTCTGTGGGGGCAAGCAGTAGTGCAGCCGGGA 431
QY 349 GACGGGCCCCGTGAGCGGCGCACACCGCGCAGTGGCGGCCCATCGTGTGTCGGCGCAAC 408
Db 432 CCACAGCTATGTATGCCCGGTGAGGAGAGTGGAGAGCAACCCCTGTATCCAGGGCA-- 489
QY 409 TACAGCCAAAGTGGAGCTGGGCGCCGACCTGGCCCTGTGCTGGCTGGCTTCAACCGCCAGC 468
Db 490 -----CGGCTCCAGGCTGACCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 539
QY 469 CTGGGCCCCCGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 528
Db 540 TTCACCAATTAATCTCTCCCGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 599
QY 529 ACCGCTGTGGGCGCACCGGCTGGGGAGAGCTCCAGAGGAGGAGATCTCTGTGCTCTCC 588
Db 600 ATGAAGTGTGGTCACTGGCTGGGCTGGGCGAGCCCTAGTGGAGAGACCTCTGCCCCGAA 659
QY 589 TGGGTGCTTACAGGAAGTGGAGCTAAGGCTGTGGCGAGGCCACCTGTCAATGTCTCTAC 648
Db 660 CGGATCTCTGAGAAACTCGCTGTGCCCATCATCGACACACCCAAAGTGCACCTGCTTTAC 719
QY 649 AGCCAGCCCGGTCCCTTCAACCTCACTCTCCAGATATT---GCCAGGAGTGTGTGTGT 705
Db 720 AGCAAGAGACCGGAGTTTGGCTACCAACCCCAACCAATCAAGATGACATGTGTGCGCC 779
QY 706 GGCTACCCAGAGGGCGCGCAGGAGACCTGCGCAGGCTGACTCTGGGGGCGCCCTGTGTGT 765
Db 780 GGCTTCGAGGAGGCGCAAGAGGATGCTGCAAGGGCGACTCGGGGCGGCCCTGTGTGTGT 839
QY 766 GAGGAAGCGGCGCCTGGTTCAGGCGAGGAATACCAAGCTTTGGGTTTGGCTGTGGACGG 825
Db 840 CTCGTGGGTCACTGTGGCTGCGAGCGGGGTGATCAGCTGGGGTGGGGCTGTGTGCGCG 899
QY 826 AGAAACCGCCCTGGAGTTTTCAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 877
Db 900 CAGNACCGCCCGAGTGTCTACATCCGTGTACCGCGGCCCAACCAACTGGATCC 951

RESULT 8
US-09-023-942A-5
; Sequence 5, Application US/09023942A
; Patent No. 6479274
; GENERAL INFORMATION:
; APPLICANT: (US only) ANTALIS Toni Marie and HOOPER John David
; TITLE OF INVENTION: NOVEL MOLECULES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 GARDEN CITY PLAZA
; CITY: GARDEN CITY
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,942A
; FILING DATE: 13-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P05101/97
; FILING DATE: 13-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP0422/97
; FILING DATE: 18-NOV-1997
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: International PCT Application
FILING DATE: 13-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, FRANK S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 11168
TELEPHONE: (516) 742 4343
TELEFAX: (516) 742 4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1100 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 17..961
US-09-023-942A-5

Query Match 14.7%; Score 161.6; DB 4; Length 1100;
Best Local Similarity 52.9%; Pred. No. 1.1e-27;
Matches 406; Conservative 0; Mismatches 344; Indels 18; Gaps 2;
QY 130 TGGGGGCGCTGAGCCCTGGCGCCGCGCATCTGGGGGGCTCAACGCGCGAGCGGCGACC 189
DB 113 TGGGGCGAGCGGTCTATCATCGTGGCGCATCGTGGGTGGAGGAGCGCGCACTCGGGCGT 172
QY 190 TGGCCCTTGGCAAGTGAAGCTGACCATGAGAGTGCCACATCTCGGGGGTCCCTCATC 249
DB 173 TGGCCCTGCGAGGAGCGCTGCGCTGTGGGATTCACAGTATGCGGAGTGAGCGTCTC 232
QY 250 GCGCCCTCTGGTCTCTCCGCTGCTACTCTTTCATGAGATGAGGAGCGTGGAGCCC 309
DB 233 AGCCACCGCTGGGCATCTACGCGCGCGCACTCTTTGAACCTTAGTACCTTAGTAT 232
QY 310 GCGGCGGAGTGGTCTGCTGCTGAGCGCTGC-----ACTCCAGGAGCGGCCCTG 360
DB 293 CCCTCCGGGTGGTGTCTCCAGTTTGGCGAGCTGACTTCCATGCGCATCTCTCGAGCGTG 352
QY 361 GAGGCGCGCACACCGCGCGAGTGGCGCCATCGTGGTGGCGGCACTACAGCAGATG 420
DB 353 CAGGCTACTACACCGCTTACTTCTGATGCAATATCTATCTGAGCGCTCGTACTCGGG 412
QY 421 GAGCTGGGCGGACCTGGCCCTGTGCGCTTGGCTCACCGCGAGCGTGGCGCCGCGC 480
DB 413 AATTACCTTATGACATGCTTGGTGGTGAAGCTGTCTGCACTGTGACCTTACCTACATAAC 472
QY 481 GTGTGGCTCTGTGCTGCGCCGCGCTCACACCGCTTGTGSCAGGCGACCGCTGCTGG 540
DB 473 ATCCAGCCCATCTGTCTCCAGGCTTCCATTTGAGTTGAGACCGGACAGACTGCTGG 532
QY 541 GCGACCGGCTGGGAGGAGCTGCGAGGAGGAGATCTCTGCTGCTCTCCCTGGTGGTCTACAG 600
DB 533 GTGACTGGCTGGGGGTACATCAAGAGGATGAGGCACTGCCATCTCCCAACCGCTCCAG 592
QY 601 GAAGTGAGCTAAGGCTGTGGCGGAGGCGACCTGTCAATGTCTTACAGCGAGCGCGGT 660
DB 593 GAACTTCAGGTGCGCATATAACAACCTATATGTGCAACCACTCTTCTCAAGTACAGT 652
QY 661 CCCCTTAACCTTACTTCCAGATATTGCGCAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTG 720
DB 653 TTCCGCAAGGAC-----ATCTTTGAGAGCATGTTTGTGCTGGCAATGCCCAAGGC 703
QY 721 CGCAGGAGCACCTGCCAGGCTGACTCTGGGGGGCGCCCTGCTGTGAGGAGGCGGCGCGC 780
DB 704 GGGAGGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 763
QY 781 TGGTCCAGGCGAGGATCACACCTTTGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 840
DB 764 TGGTATCAGATTGGAGTCGTGAGCTGGGAGTGGGCTGTGGTGGGCGCCCAATCGGCGCGGT 823

QY 841 GTTTTCACCTGCTGTGGCTTACCTATGAGCGCATGGATATCGGAGCGAGGTG 888
DB 824 GTCTACCAATATCAGCCACCACCTTTTGGTGGATCCAGAAAGCTGATG 871
RESULT 9
US-09-907-794A-256
Sequence 256, Application US/09907794A
Patent No. 6635468
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,794A
CURRENT FILING DATE: 2001-07-17
PRIORITY APPLICATION NUMBER: PCT/US00/04414
PRIORITY FILING DATE: 2000-02-22
PRIORITY APPLICATION NUMBER: US 60/143,048
PRIORITY FILING DATE: 1999-07-07
PRIORITY APPLICATION NUMBER: US 60/145,698
PRIORITY FILING DATE: 1999-07-26
PRIORITY APPLICATION NUMBER: US 60/146,222
PRIORITY FILING DATE: 1999-07-28
PRIORITY APPLICATION NUMBER: PCT/US99/20594
PRIORITY FILING DATE: 1999-09-08
PRIORITY APPLICATION NUMBER: PCT/US99/20944
PRIORITY FILING DATE: 1999-09-13
PRIORITY APPLICATION NUMBER: PCT/US99/21090
PRIORITY FILING DATE: 1999-09-15
PRIORITY APPLICATION NUMBER: PCT/US99/21547
PRIORITY FILING DATE: 1999-09-15
PRIORITY APPLICATION NUMBER: PCT/US99/23089
PRIORITY FILING DATE: 1999-10-05
PRIORITY APPLICATION NUMBER: PCT/US99/28214
PRIORITY FILING DATE: 1999-11-29
PRIORITY APPLICATION NUMBER: PCT/US99/28313
PRIORITY FILING DATE: 1999-11-30
PRIORITY APPLICATION NUMBER: PCT/US99/28564
PRIORITY FILING DATE: 1999-12-02
PRIORITY APPLICATION NUMBER: PCT/US99/28565
PRIORITY FILING DATE: 1999-12-02
PRIORITY APPLICATION NUMBER: PCT/US99/30095
PRIORITY FILING DATE: 1999-12-16
PRIORITY APPLICATION NUMBER: PCT/US99/30911
PRIORITY FILING DATE: 1999-12-20
PRIORITY APPLICATION NUMBER: PCT/US99/30999
PRIORITY FILING DATE: 1999-12-20

/	PRIOR APPLICATION NUMBER:	PCT/US00/00219			
/	PRIOR FILING DATE:	2000-01-05			
/	NUMBER OF SEQ ID NOS:	423			
/	SEQ ID NO	256			
/	LENGTH:	1100			
/	TYPE:	DNA			
/	ORGANISM:	Homo Sapien			
/	US--09-907-794A-256				
	Query Match	14.7%; Score 161.6; DB 4; Length 1100;			
	Best Local Similarity	52.9%; Pred. No. 1.1e-27;			
	Matches	406; Conservative 0; Mismatches 344; Indels 18; Gaps 2;			
QY	130	TGCGGGCCCTGAGCCCTCGGCCGCATCGTGGGGGGCTCAACGCGCAGCGCGGACC 189			
DB	116	TGCGGGCGACGGGTCACTACGTCCGCATTCGTGGTGAGAGAGCGCGACTCGGGCGT 175			
QY	190	TGGCCTTTGGCAAGTAGCCTGCACATGAGAGTGGCCACATCTGCGGGGCTCCCTCATC 249			
DB	176	TGGCCGTGGCAGGGAGCCTGGCCTGTGGATTCCCACGTATGCGAGTGAAGCTGCTC 235			
QY	250	GCCCCCTCTGGGTCTCTCCGTGCTCACTGTTCATGACGATGGAGCGTTGGAGCCC 309			
DB	236	AGCCACCGCTGGGCACCTCACGGCGCGCACTGCTTTGAAACCTATAGTAGACCTTAGT 295			
QY	310	CGCGCCAGTGGTCCGTACTCTCTGGGCGTGC-----ACTCCAGGACGGGCCCCCTG 360			
DB	296	CCCTCCGGGTGGATGGTCCAGTTTGGCCAGCTGACTTCATGCCATCCTTCTGGAGCCTG 355			
QY	361	GACGGCGGCACACCGCGCAGTGGCGGCCCATCGTGTGCGGCCAACTACAGCCAAGTG 420			
DB	356	CAGGCTTACTACA CCGGTTACTTCGTATCGAATACTATCTGAGCCCTGCTACCTGGGG 415			
QY	421	GAGCTGGCGCGCACTGGCCCTGTCGGCCTGAGCCTCACCCGCGAGCTGGGCCCCGCC 480			
DB	416	AATTCAACCCTATGACATTGCTTGGTGAAGCTGTCTGCACCTGTCACTACACTAAACAC 475			
QY	491	GTGTGGCTGTCTGCTGCCCGCGCCTCACACCGTCTCGTGCAGCGCACCGCTGCTGG 540			
DB	476	ATCCAGCCCATCTGTCTCCAGCCCTCCCAATTGAGTTTGAGAACCGGACAGACTGTGG 535			
QY	541	GCCACGGCTGGGAGACGCTCCAGGAGGAGATCCTCTGCGCTCTCCCCCTGGGTGCTACAG 600			
DB	536	GTGACTTGGCTTGGGGTAATCAAAGAGATGAGGCACCTGCCATCTCCACACCCCTCCAG 595			
QY	601	GAAGTGGAGCTAAGGCTGCTGGCGGAGGCCA CCTGTCAATGTCTCTACAGCAGCCCCGT 660			
DB	596	GAA GTTCAGGTGCGCCATCAATAAACACTCTATGTGCAACCACTCTTCTCAAGTACAGT 655			
QY	661	CCCTTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGTGTGGCTTACCACAGAGGC 720			
DB	656	TTCCGCAAGGAC-----ATCTTTGGAGACATGTTTGTGTGGCAACGCCCAAGGC 706			
QY	721	CGCAGGACACCTGCCAGGTGACTCTGGGGGCCCTGCTGTGTGAGGAAGCGCGCGCCG 780			
DB	707	GGGAAGAATGCCCTGCTTGGTGTACTCAGTTGGACCTTGGCCCTGTAAACAAGAAATGGACTG 766			
QY	781	TGTTTCCAGGCAGGAATCA CCAAGCTTTGGTTTTGGCTGTGGACGGAGAAACCGCCCTGGA 840			
DB	767	TGCTATCAGATTGAGTCTGTGACTGGGGAGTGGGCTGTGGTCGGGCCCAATCGGCCCGGT 826			
QY	841	GT TTTTCACTGCTGGCTACTATTAGGCATGATACGGGAGCAGCTG 888			
DB	827	GTCATACCAATATACAGCCACCACTTTGAGTGGATCCAGAAGCTGATG 874			

```

RESULT 10
US/09-905-125A-256
; Sequence 256, Application US/09905125A
; Patent No. 6664376
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi

```

```

1  APPLICANT: Botstein, David
2  APPLICANT: Desnoyers, Luc
3  APPLICANT: Eaton, Dan L.
4  APPLICANT: Ferrara, Napoleone
5  APPLICANT: Filvaroff, Ellen
6  APPLICANT: Fong, Sherman
7  APPLICANT: Gao, Wei-Qiang
8  APPLICANT: Gerber, Hanspeter
9  APPLICANT: Gerritsen, Mary E.
10 APPLICANT: Goddard, A.
11 APPLICANT: Godowski, Paul J.
12 APPLICANT: Grimaldi, Christopher J.
13 APPLICANT: Gurney, Austin L.
14 APPLICANT: Hillan, Kenneth, J.
15 APPLICANT: Kljavin, Ivar J.
16 APPLICANT: Macher, Jennie P.
17 APPLICANT: Pan, James
18 APPLICANT: Paoni, Nicholas F.
19 APPLICANT: Roy, Margaret Ann
20 APPLICANT: Stewart, Timothy A.
21 APPLICANT: Tumas, Daniel
22 APPLICANT: Williams, P. Mickey
23 APPLICANT: Wood, William, I.
24
25 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
26
27 FILE REFERENCE: 10466-14
28
29 CURRENT APPLICATION NUMBER: US 09/905.125A
30
31 CURRENT FILING DATE: 2001-07-12
32
33 PRIOR APPLICATION NUMBER: PCT/US00/04414
34
35 PRIOR FILING DATE: 2000-02-22
36
37 PRIOR APPLICATION NUMBER: US 60/143,048
38
39 PRIOR FILING DATE: 1999-07-07
40
41 PRIOR APPLICATION NUMBER: US 60/145,698
42
43 PRIOR FILING DATE: 1999-07-26
44
45 PRIOR APPLICATION NUMBER: US 60/146,222
46
47 PRIOR FILING DATE: 1999-07-28
48
49 PRIOR APPLICATION NUMBER: PCT/US99/20594
50
51 PRIOR FILING DATE: 1999-09-08
52
53 PRIOR APPLICATION NUMBER: PCT/US99/20944
54
55 PRIOR FILING DATE: 1999-09-13
56
57 PRIOR APPLICATION NUMBER: PCT/US99/21090
58
59 PRIOR FILING DATE: 1999-09-15
60
61 PRIOR APPLICATION NUMBER: PCT/US99/21547
62
63 PRIOR FILING DATE: 1999-09-15
64
65 PRIOR APPLICATION NUMBER: PCT/US99/23089
66
67 PRIOR FILING DATE: 1999-10-05
68
69 PRIOR APPLICATION NUMBER: PCT/US99/28214
70
71 PRIOR FILING DATE: 1999-11-29
72
73 PRIOR APPLICATION NUMBER: PCT/US99/28313
74
75 PRIOR FILING DATE: 1999-11-30
76
77 PRIOR APPLICATION NUMBER: PCT/US99/28564
78
79 PRIOR FILING DATE: 1999-12-02
80
81 PRIOR APPLICATION NUMBER: PCT/US99/28565
82
83 PRIOR FILING DATE: 1999-12-02
84
85 PRIOR APPLICATION NUMBER: PCT/US99/30095
86
87 PRIOR FILING DATE: 1999-12-16
88
89 PRIOR APPLICATION NUMBER: PCT/US99/30911
90
91 PRIOR FILING DATE: 1999-12-20
92
93 PRIOR APPLICATION NUMBER: PCT/US99/30999
94
95 PRIOR FILING DATE: 1999-12-20
96
97 PRIOR APPLICATION NUMBER: PCT/US00/00219
98
99 PRIOR FILING DATE: 2000-01-05
100
101 NUMBER OF SEQ ID NOS: 423
102
103 SEQ ID NO 256
104
105 LENGTH: 1100
106
107 TYPE: DNA
108
109 ORGANISM: Homo Sapien
110
111 US-09-905-125A-256

```

Query Match	14.7%;	Score 161.6;	DB 4;	Length 1100;
Best Local Similarity	52.9%;	Pred. No. 1.1e-27;		
Matches 406; Conservative		0; Mismatches 344;	Indels 18;	Gaps 2;

Qy	130	TGCGGGCGCCCTGAGCCCTCGGCCCGATCGTGGGGGGCTCAAACGCGAGACCGGGCAC	189
Db	116	TGCGGCCGACCGGGTCATCACGTCGCGCATCGTGGGTGGAGAGACCGCGAACTCGGGGGT	175
Qy	190	TGGCCTTGGCAAGTGAGGCTGCACCATGAGAGTGGCCACATCTGCGGGGGCTCCCTCATC	249
Db	176	TGGCCGTGGCAGGGGAGGCTGGCCCTGTGGGATTCCTCAGTATGGGAGTGAAGCTGTCT	235
Qy	250	GCCCCCTCTGGGTCTCTCCGCTGCTCACTGTTCATGAGAAATGGGACGTTGAGGCC	309
Db	236	AGCCACCGCTGGGCACTCACGGCGCGCACTGCTTTGAAACCTATAGTGACCTTAGTGAT	295
Qy	310	GCGGCGAGTGGTGGTACTGCTGGGGCTGC-----ACTCCAGAGACGGGGCCCTG	360
Db	296	CCCTCCGGGTGGATGGTCCAGTTGGCCAGCTGACTTCATGCCATCTCTCTGAGCCCTG	355
Qy	361	GACGGCGGCACACCGCGAGTGGCCGCCATCGGTGTGCGGGCCAACTACAGGCCAAATG	420
Db	356	CAGGCTACTAACCCGTTACTTGTATCGAATATCTATCTGAGCCCTGCTGCTTGGGG	415
Qy	421	GAGCTGGCGCGCAGCTTGGCCCTGTGCGCCCTGGGCTTCAACCGCCAGAGCTGGGGCCCGC	480
Db	416	AATTCAACCTATGATTTGCTTGGTGAAGTGCTGTGCACCTGTCACTTACACTAAACAC	475
Qy	481	GTGTGGCTGTCTGGCTGCCCGGCGCTCAACCGCTTCTGTGCAAGGACACCGCTGCTGG	540
Db	476	ATCAGGCCCACTGTCTGTCCAGCCCTCCACATTTGAGTTTGAAGACGGACAGACTGCTGG	535
Qy	541	GCCACCGCTGGGAGAGAGTCCAGAGGACAGATCCTCTGCCCTCTCCCTGGGTGCTACAG	600
Db	536	GTGACTGGCTGGGGGTAATCAAGAGGATGAGGCACCTGCCAATCTCCCAACCCCTCCAG	595
Qy	601	GAACTGGAGCTAAGGCTGCTGGGCGAGGCGCACTGTCAATGTCTCTACAGCCAGACCCGGT	660
Db	596	GAACTTCAGGTGCGCATCATATAACAACTCTATGTGCAACCACTCTTCTCAAGTACAGT	655
Qy	661	CCCTTCAACCTCACTTCCAGATATGCCAGGATGCTGTGTGTGCTGGCTTACCCAGAGGC	720
Db	656	TTCCGCAAGGAC-----ATCTTTGAGACATGGTTTGTGTGTGGCAAGCCCAAGGC	706
Qy	721	CGCAGGACACTGCCAGGTGACTCTGGGGGGCCCTGGTCTGTGTGAGGAAGGGCGCGC	780
Db	707	GGGAAGGATGCTGCTTCGGTGTACTCAGGTGGACCCCTTGGCTGTAAACAAGAAATGGACTG	766
Qy	781	TGGTTCCAGGACGGAATCACACAGCTTTTGGGTTTGGCTGTGGACGGAGAAACCGCCCTGA	840
Db	767	TGGTATCAGATTGGAGTGTGAGCTGGGAGTGGGCTGTGGTTCGCCCCATCGGCCCGT	826
Qy	841	GTTTTCACTGCTGTGGCTACCTATGAGGCAATGGATACGGGAGCAGGTG	888
Db	827	GTCTACACCAATATCAGCACCACTTTTGTGTGATCCAGAACTGATG	874

```

RESULT 11
US-09-902-775A-256
; Sequence 256, Application US/09902775A
; Patent No. 6686451
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.

; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-902-775A-256

Query Match      14.7%; Score 161.6; DB 4; Length 1100;
Best Local Similarity 52.9%; Pred. No. 1.1e-27;
Matches 406; Conservative 0; Mismatches 344; Indels 18; Gaps 2;

Qy 130 TCGCGCGCCCTGAGCGCTCGGCCGCGCATCGTGGGGGGCTCAAAAGCGGACGCGGGCACC 189
      |||||
Db 116 TCGGGCCGACGGGTATCATCAGCTCGCGCATCGTGGGTGGAGAGGACGCCGAATCTCGGGCGT 175

Qy 190 TGGCCTTGGCAAGTGAGCCTGCAACATGGAGGTGGCCACATCTCGGGGGGTCCCTCATC 249
      |||||
Db 176 TGGCCCGTGGCAGGGAGGCGCTGGCGCTGTGGGATTTCCCAACGATATCGGAGTGAGCGTCTC 235

Qy 250 GCCCCCTCCTGGGTCTCTCGCGTGTCTCACATGTTTTCATGACGAATGGAGCGTTGGAGCCC 309
      |||||
Db 236 AGCCACCGCTGGGCGACTCACGGCGGCGACCTGCTTTGAACCTTATAGTGCATTAGTGAT 295

Qy 310 GCGGCGAGTGGTGGGTACTGCTGGGGCGTGC-----ACTCCCGAGGAGGGGCCCTCG 360

```


RESULT 13

US-09-023-942A-25
; Sequence 25, Application US/09023942A
; Patent No. 6479274
; GENERAL INFORMATION:
; APPLICANT: (US only) ANTALIS Toni Marie and HOOPER John David
; TITLE OF INVENTION: NOVEL MOLECULES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 GARDEN CITY PLAZA
; CITY: GARDEN CITY
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,942A
; FILING DATE: 13-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: POS101/97
; FILING DATE: 13-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP0422/97
; FILING DATE: 18-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: International PCT Application
; FILING DATE: 13-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: DIGIGLIO, FRANK S
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 11168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742 4343
; TELEFAX: (516) 742 4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 959 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..856
US-09-023-942A-25

Query Match 14.3%; Score 157.6; DB 4; Length 959;
Best Local Similarity 52.4%; Pred. No. 8.8e-27;
Matches 377; Conservative 0; Mismatches 334; Indels 9; Gaps 1;
QY 129 CTGGGGGCGCCCTGAGCCCTGGCCCGCATCGTGGGGGGCTCAGACGCGCGCGGCAC 188
Db 19 CTGGCGGTACAGACATCCCTTCCGTATGTTGGTGGCGATGATGCTGAGCTTGGCGG 78
QY 189 CTGGCCCTTTGGCAAGTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 248
Db 79 CTGGCGGTGGCAAGGAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 138
QY 249 CGCCCGCTCTGGGCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 308
Db 139 CAACCGCGCTGGGCTTACAGCTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 198
QY 309 CGCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 368

Db 199 CTGGACAGTCCAGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 258
QY 369 GCACACCGCGCAGTGGCCCGCCATCGTGGTGGCGGCCCACTACAGCCCAAGTGGAGCTGGG 428
Db 259 TTCCACACCGTTACCAATAGAGATATTTTCTGAGCCCCCAAGTACTCGGAGCAGTATCC 318
QY 429 CGCGACCTGGCCCTGCTGGCGCTGGCTCACCGCCGCGCAGCTGGGCCCGCCCGCTGGTGG 488
Db 319 CAATGACATAGCCCTGCTGAAGCTGTCTATCTCCAGTCACTACATAAATTTTATCCAGCC 378
QY 489 TGCTGCTGCGCCCGCCCTCACACCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 548
Db 379 CATCTGCTCTCTGAACCTCCACGCTACAAAGTTTGAAGACCGAACTGACTGCTGGTGG 438
QY 549 CTGGGGAGAGCTCCAGGAGGAGATCTCTGCTGCTCTCCCTGGTGGTGGTGGTGGTGGTGG 608
Db 439 CTGGGGGGCTATTGGAGAGATGAGAGTCTGCCATCTCCCAACACTCTCCAGGAGTGA 498
QY 609 GCTAAGGCTCTGGCGGAGGCCACTGTCTCAATGTCTACAGCCGCGCCCGCTGGTGGTGG 668
Db 499 GGTAGCTATTATCAACAACAGCATGTCTAACCATATGTACAAAAAGCCAG-----A 549
QY 669 CTTCACTCTCCAGATATTGCCAGGATGCCAGGATGCTGTGCTGGTGGTGGTGGTGGTGG 728
Db 550 CTTCCGCAACAACATCTGGGGAGACATGGTTGGCTGGGCACTCTGAAAGTGGCAAGGA 609
QY 729 CACCTGCAGGCTGACCTCTGGGGGGCCCTGGTCTGTGAGGAAGCGCGCGCTGGTTCCA 788
Db 610 TGCTGCTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 669
QY 789 GGCAGGAATCAGCAGCTTTGGGTTGGCTGTGGACGGAGAAACCGCCCTGGAGTTTTCAC 848
Db 670 GGTGGAGTTGTGAGCTGGGGGAATAGGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 729

RESULT 14

US-09-023-942A-3
; Sequence 3, Application US/09023942A
; Patent No. 6479274
; GENERAL INFORMATION:
; APPLICANT: (US only) ANTALIS Toni Marie and HOOPER John David
; TITLE OF INVENTION: NOVEL MOLECULES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 GARDEN CITY PLAZA
; CITY: GARDEN CITY
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,942A
; FILING DATE: 13-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: POS101/97
; FILING DATE: 13-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP0422/97
; FILING DATE: 18-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: International PCT Application
; FILING DATE: 13-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: DIGIGLIO, FRANK S
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 11168
; TELECOMMUNICATION INFORMATION:

```

, , TELEPHONE: (516) 742 4343
, , TELEFAX: (516) 742 4366
, , TELEX: 230 901 SANS UR
, , INFORMATION FOR SEQ ID NO: 3:
, , SEQUENCE CHARACTERISTICS:
, , LENGTH: 1094 base pairs
, , TYPE: nucleic acid
, , STRANDEDNESS: single
, , TOPOLOGY: linear
, , MOLECULE TYPE: DNA
, , FEATURE:
, , NAME/KEY: CDS
, , LOCATION: 17..955
, , US-0923-942A-3

```

Query Match.	14.0%;	Score 154.8;	DB 4;	Length 1094;
Best local Similarity	52.2%;	Pred. 3.9e-26;		
Matches	398;	Conservative 0;	Mismatches 352;	Indels 12; Gaps 2;
QY	130	TGCGGGCGCCTGAGCCCTCGGCCGCGCATCGTGGGGGGCTCAACAGCGCAGCGCGGCACC	189	
DB	113			
QY	190	TGGCCTTTGGCAAGTGAGCCTGCACATATGAGAGTGGCCACATCTCGGGGGGTCTCCTCATC	249	
DB	173	TGGCCGTGGCAGGGAGCCTCGCCCTGTGGATTTCCACGATGCGGAGTGAGCTGCTCTC	232	
QY	250	GC CCCTCTCGGGTCTCTCGCGTCTCAGTGTTCATGACGATGGAGCGTGGAGGCC	309	
DB	233	AGCCACCGCTGGGCACCTCA CGCGCGCGACCTGTCTTGAACTGACCTTATGTATGATCCCTCC	292	
QY	310	G---CGGCGAAGTGGTCGTACTGCTGGCGGTGCACCTCCACGAGCGGGCCCTCGAGCGGC	366	
DB	293	GGGTGGATGGTTCAGTTTGGCCAGCTGATTCATGCCATCTCTTCTTGAGGCTTCAGAGCC	352	
QY	367	GC GCACACCCCGCAGTGGCGGCCCATCGTGGTGGCGGCCAACTACGCAAGTGGAGGTG	426	
DB	353	TACTACACCCGTTACTTCGTATCGAATATCTATCTGAGCCCTCGTACTCTGGGAATTC	412	
QY	427	GGCGCGACCTGGCGCCTGCTGGCGCTTGGCCTCACCGCGCAGCTTGGGCGCCGCGGTGG	486	
DB	413	CCCTATGACATTTGCTTGGTGAAGCTGTCTGACCTGTCTACCTACCTAACAACATCCAG	472	
QY	487	CCTGTCTGCTTGC CGCGCGCCTCACACCGCTTCGTGACGGCACC CGCTGCTGGGGCCACC	546	
DB	473	CCCATCTGCTCCAGGCCTCCACATTTGAGTTTGAGAACCGGACAGACTGCTGGGTGACT	532	
QY	547	GGCTGGGAGAGCTCCAGGAGGCGAGATCCTCTG CCCTCTCCCGTGGGTGCTACAGGAAGTG	606	
DB	533	GGCTGGGGGTACATCAAAAGAGATGAGGCACCTGCCATCTCCCCACACCGCTCCAGGAAGTT	592	
QY	607	GAGCTAAGGCTGCTGGGCGAGGCCACCHGTCAA TGTCTCTACAGCAGCCCGTCCCTTC	666	
DB	593	CAGTTCGCCATCATAAACAATCTATGTGCAACCACTTCTCTCAAGTACAGTTTCGCG	652	
QY	667	AACCTCACTCTCCAGATATTGCCAGGGATGCTGTGTGTGCTATCCAGAGGGCGCGCAGG	726	
DB	653	AAGGAC-----ATCTTTGGAGACATGTTTGTGTGGCAATGCCAAGGCGGGGAAG	703	
QY	727	GACACCTGCCAGGTGTACTTGGGGGGGCCCTGTGTC GTGTGAGGAAGCGCGCGCTGTGTC	786	
DB	704	GATGCTCTGCTTCGTGTGACTCAGGTGGACCCCTTGGCCTGTAA CAGGATGTGGTAT	763	
QY	787	CAGGCAGGAATCACCAGCTTTTGGCTTGGCTGTGACGAGGAGAAACCGCCCTGGAGTTTTC	846	
DB	764	CAGATTGAGTCTGAGCTGTGGAGTGTGGCTGTGCTCGGCCCAATCGGCCCGGTCTCTAC	823	
QY	847	ACTGTGTGGCTACTATGAGGCATGGATPACGGGAGCAGGTG	898	
DB	824	ACCAATATCAGCCACCATTTTCAGTGGATTCACAGAGCTGATG	865	

RESULT 15

US-09-386-653A-8
; Sequence 8, Application US/09386653A
; Patent No. 6458564
; GENERAL INFORMATION:
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jian-shen
; TITLE OF INVENTION: DNA encoding the novel human serine
; TITLE OF INVENTION: protease T
; FILE REFERENCE: ORT-1032
; CURRENT APPLICATION NUMBER: US/09/386,653A
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1130
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene of
; OTHER INFORMATION: Protease T in a zymogen activation vector
US-09-386-653A-8

Query Match 14.0%; Score 153.8; DB 4; Length 1130;
Best Local Similarity 53.9%; Pred.No. 6.6e-26;
Matches 392; Conservative 0; Mismatches 317; Indels 18; Gaps 3;

QY	157	ATCGTGGGGGCTCAACGCGCAGCGGCGACCTTGGCCTTTGGCAAGTAGGCTGCACCAT	216
DB	166	ATCGTGGGGGCTATGCTCTAGAGGAGGGCGAGTGGCCCTTGGCAAGTACGATCCAGGCG	225
QY	217	GGAGGTGGCCACATCTCGGGGGGCTCCCTCATCGCCCCCTCTCGGGTCTCTCCCGTGGCT	276
DB	226	AACGGAAGCCACTTCTCGGGGGGCGAGCCTCATCGCGAGCAGTGGGTCTCTGACGGCTGCG	285
QY	277	CACGTGTTT---CATGACGAATGGACAGTGGAGCCCGCGGCCGAGTGGTGGTACTGCTG	333
DB	286	CACGTGCTTCGCGACACACTCTGAGACCTCCCTGTACCAGGTCTCTGCTGGGGGCAAGGAG	345
QY	334	GGCGTGCACATCCAGGACGGGCCCTCGACGGCGCGCACACCGCGCAGGTGGCGCGCCATC	393
DB	346	CTAGTGACGGGGACACACGCTATGTATGCCGGGTGAGGAGGTGAGAGCAACCC	405
QY	394	GTGTGCGCGGCAACATACAGCAAGTGGAGCTGGGGCGGACCTGGGCCCTGCTGGGCTG	453
DB	406	CTGTACCAAGGCA-----CGGCCCTCCAGCGCTGACGTGGGCCCTGGTGGAGCTG	453
QY	454	GCCTCACCGCGCAGCTGGGCCCGCGCTGTGGCTGTCTCTCTGCCCGCGGCTCCACAC	513
DB	454	GAGGCAACAGTGCCCTTACCAATATCATCTCTCCCGTGGCTGCTGACCCCTCGGTG	513
QY	514	CGCTTCGTGCACGGCACCGCGCTGTGGGCCACCGGCTGGGGAGACGTCCAGAGGCGAGAT	573
DB	514	ATCTTTGAGACGGGCATGAACCTGTGGGTCACTGGCTGGGGCAGCCCCAGTGAGGAAGAC	573
QY	574	CCTCTGGCTCTCCCGTGGGTGCTACAGGAAGTGGAGCTTAAGGCTGCTGGGGAGGCCACC	633
DB	574	CTCCTGCCCGAACCCGCGATCCTCGAAACTCGCTGTGCCCATCATCGACACACCCCAAG	633
QY	634	TGTCAAATGCTCTACAGCCAGCCCGGTCCTTTCAACCTCACTCTCCAGATATT---GCCA	690
DB	634	TGCAACCTGCTCTACAGCAAGACACCGAGTTTGGCTACCAACCCAAAACCATCAAGAT	693
QY	691	GGGATGTGTGTGTGCTTACCCAGAGGGCGGACAGGACACCTCCAGGTGATCTTGGG	750
DB	694	GACATGCTGTGCGCGGGCTTCGAGGAGGGCAAGAGGATGCTCTCAAGGGGAGCTCGGGC	753
QY	751	GGGCCCTGTGTGTGAGGAAGGCGCGCTGGTTCCAGGCAGGAATCACAGCTTTGGG	810
DB	754	GGCCCCCTGGTGTGCTCTGTGGGTCACTGTGGCTGTGACGGCGGGGTGATCAGCTGGGT	813
QY	811	TTTGGCTGTGACGGAGAAACCGGCCCTGGAGTTTTCATGCTGTGGGTACTATGAGGCA	870

Db B14 GAGGGCTGTGCCCGCCAGAACCGCCAGGTGTCTACATCGGTGTACCGGGCCCAAC 873
QY 871 TGGATAC 877
Db 874 TGGATCC 880

Search completed: February 25, 2004, 11:19:47
Job time : 118 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 10:31:35 ; Search time 452 Seconds
(without alignments)
8537.195 Million cell updates/sec

Title: US-10-037-417-45
Perfect score: 1102
Sequence: 1 gggccctgtctggccat.....ggggtctgatggggcctccc 1102

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2308694 seqs, 1750822206 residues 4617368

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq.*
- 17: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq.*
- 18: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	815.6	74.0 2457 9	US-09-888-615-52 Sequence 52, Appl
2	232.4	21.1 3382 14	US-10-101-510-447 Sequence 447, Appl
3	230.8	20.9 1733 14	US-10-176-847-85 Sequence 85, Appl
4	225.6	20.5 1020 15	US-10-051-874-25 Sequence 25, Appl
5	224	20.3 1613 13	US-10-041-400A-1 Sequence 1, Appl
6	224	20.3 1613 13	US-10-041-264A-1 Sequence 1, Appl
7	224	20.3 1613 13	US-10-042-091A-1 Sequence 1, Appl
8	221.2	20.1 1834 9	US-09-948-094-1 Sequence 1, Appl
9	221.2	20.1 1834 9	US-09-880-107-2214 Sequence 2214, Ap
10	221.2	20.1 1834 9	US-09-967-768A-141 Sequence 141, App
11	221.2	20.1 1834 14	US-10-097-340-261 Sequence 261, App
12	220.6	20.0 1668 9	US-09-925-301-208 Sequence 8, Appl
13	216.8	19.7 1130 13	US-10-041-400A-8 Sequence 8, Appl
14	216.8	19.7 1130 13	US-10-041-264A-8 Sequence 8, Appl
15	216.8	19.7 1130 13	US-10-042-091A-8 Sequence 8, Appl

16	195.2	17.7	1797	14	US-10-109-616-1	Sequence 1, Appl
17	185.6	16.8	882	12	US-10-042-865-33	Sequence 33, Appl
18	184	16.7	882	12	US-10-042-865-34	Sequence 34, Appl
19	180.8	16.4	768	14	US-10-221-097-10	Sequence 10, Appl
20	179	16.2	1161	12	US-10-042-865-31	Sequence 31, Appl
21	176.2	16.0	1327	9	US-09-978-295A-170	Sequence 170, App
22	176.2	16.0	1327	9	US-09-978-697-170	Sequence 170, App
23	176.2	16.0	1327	9	US-09-978-192A-170	Sequence 170, App
24	176.2	16.0	1327	9	US-09-999-832A-170	Sequence 170, App
25	176.2	16.0	1327	10	US-09-978-189-170	Sequence 170, App
26	176.2	16.0	1327	10	US-09-978-608A-170	Sequence 170, App
27	176.2	16.0	1327	10	US-09-978-585A-170	Sequence 170, App
28	176.2	16.0	1327	10	US-09-978-191A-170	Sequence 170, App
29	176.2	16.0	1327	10	US-09-978-403A-170	Sequence 170, App
30	176.2	16.0	1327	10	US-09-978-564A-170	Sequence 170, App
31	176.2	16.0	1327	10	US-09-999-833A-170	Sequence 170, App
32	176.2	16.0	1327	10	US-09-981-915A-170	Sequence 170, App
33	176.2	16.0	1327	10	US-09-978-824-170	Sequence 170, App
34	176.2	16.0	1327	10	US-09-918-585A-170	Sequence 170, App
35	176.2	16.0	1327	10	US-09-978-423A-170	Sequence 170, App
36	176.2	16.0	1327	10	US-09-978-193A-170	Sequence 170, App
37	176.2	16.0	1327	10	US-09-999-830A-170	Sequence 170, App
38	176.2	16.0	1327	10	US-09-978-757A-170	Sequence 170, App
39	176.2	16.0	1327	10	US-09-978-187B-170	Sequence 170, App
40	176.2	16.0	1327	10	US-09-978-643A-170	Sequence 170, App
41	176.2	16.0	1327	10	US-09-978-375A-170	Sequence 170, App
42	176.2	16.0	1327	10	US-09-978-298A-170	Sequence 170, App
43	176.2	16.0	1327	10	US-09-978-188A-170	Sequence 170, App
44	176.2	16.0	1327	10	US-09-978-681A-170	Sequence 170, App
45	176.2	16.0	1327	10	US-09-978-194A-170	Sequence 170, App

ALIGNMENTS

RESULT 1
US-09-888-615-52
; Sequence 52, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CHARENPCEK, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-888-615-52

Query Match				74.0%;	Score 815.6;	DB 9;	Length 2457;
Best Local Similarity				99.5%;	Pred. No. 1.3e-204;		
Matches 818;				Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	124	CGCTACTCGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAACCGCGACGCCG	183				
Db	106	CTGGATCTGGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAACCGCGACGCCG	165				
QY	184	GGCACCCTGGCTTGGCAAGTGAGCCCTGCAACCATGAGGTGGCCACATCTCGCGGGGGCTCC	243				
Db	166	GGCACCCTGGCTTGGCAAGTGAGCCCTGCAACCATGAGGTGGCCACATCTCGCGGGGGCTCC	225				
QY	244	CTCATCGCCCCCTCTGGGTCTCTCCGCTGCTCCTGTTTCATGACGAATGGGACGTTCG	303				


```

; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 85
; LENGTH: 1733
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-176-847-85

```

```

Query Match      20.9%; Score 230.8; DB 14; Length 1733;
Best Local Similarity 54.9%; Pred. No. 5.9e-51;
Matches 507; Conservative 0; Mismatches 402; Indels 15; Gaps 2;

Qy 1 GGGCCCTTGTCTGGCCCATGGCCAGAGGGGTCTCTGGGGCTTGGGCAGCTGGGGCT 60
Db 81 GGGCCCTTGTCTGGCCCATGGCCAGAGGGGTCTCTGGGGCTTGGGCAGCTGGGGCT 140
Qy 61 GTGGCCATTCTGACTCATCTCACTTTACCGGTTTGGTGCCTCCGGACCCGCTAGGGGC 120
Db 141 GTGGCCATTCTGACTCATCTCACTTTACCGGTTTGGTGCCTCCGGACCCGCTAGGGGC 200
Qy 121 CCCCGCTACTGGGGGCGCCCTGAGCCCTCGGCCGCAATCGTGGGGGCTCAAAACGCGCAG 180
Db 201 GCTCCCTCGGTGTG-----GCCCGCCCAAGCAGCATCACAGGTGGCAGCATGCGATC 254
Qy 181 CGGGCAGCTGGCTTGGCAAGTGAAGCTGACCATGACCATGAGGTGGCAGCATTGGGGGGC 240
Db 255 GCGGTCAGTGGCCCTGGCAGGTGACGATGACCATGATGAAGGCGTCCATGTGTGTGGTGGC 314
Qy 241 TCCCTCATGCCCCCTCTCGGTCTCTCTCGGTCTCTCTCGGTCTCTCTCATGCAATGGGACG 300
Db 315 TCTCTCGTGTGTGACGAGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 367
Qy 301 TTGGAGCCCGGGCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
Db 368 --GCACACAGAGAGCCTATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 425
Qy 361 GACGGGGGCGCACACCGGGGAGTGGCGGCTATGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
Db 426 GAGGAGCGCGAAGGTGAGCACCTTGAAGGACATCATCCCCCAGCCAGCTTACCTCCAGGAG 485
Qy 421 GAGCTGGGGCGGAGCTGCGCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 480
Db 486 GGTCTCGAGGGGCGATTTGATCTCTTCCNACTCAGCAGACCCATCACCTTCTCCGCTAC 545
Qy 481 GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db 546 ATCCGGCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 605
Qy 541 GGCACGGGTGGGAGAGCTCAGGAGGAGATCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 600
Db 606 GTCACTGGCTGGGGTCTATGTGCCCCCTCAGTAGAGCTCTGAGGCCCAAGCCATCTGAG 665
Qy 601 GAAGTGGAGCTAAGGTCTGTGGCGAGGCGACCTGTCAATGTCTTACAGCCAGCCCGGT 660
Db 666 CAACTCGAGGTGCTCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 725
Qy 661 CCCTTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGTGTGTGTGTGTGTGTGTGT 720
Db 726 AAGCCTGAGAGCGGACCTTTGTCAGAGGACATGCTGTGTGTGTGTGTGTGTGTGTGTGT 785
Qy 721 CGCAGGAGACCTGCGCAGGATGCTGTGGGGGGCCCTGCTGTGTGTGTGTGTGTGTGTGT 780
Db 786 GGCAGAGCGCCTTGCAGGGGTGATCTGTGGGGGGCCCTCTCTCTCTCTCTCTCTCTCTCT 845
Qy 781 TGGTTCAGGACAGGATCACACGCTTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840
Db 846 TGGTACTGACGGGATGTGTGAGCTGGGGAGATGCTGTGTGGGGCCCGCAGAGGCTGTGT 905
Qy 841 GTTTTCACTGCTGTGGCTACCTATGAGGCATGATGAGGAGGAGGATGAGGATGAGGTT 900

```

```

Db 906 GTGTACACTCTGGCTCCAGCTATGCTCTGTGATCCAAAGCAAGGTGACAGAACTCCAG 965
Qy 901 CCTGGGCTGCTTCCACCCAG 924
Db 966 CCTCGTGTGTGCTCCCAAAACCCAG 989

RESULT 4
US-10-051-874-25
; Sequence 25, Application US/10051874
; Publication No. US20040005557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldos, Ferenc AM
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Macdougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol BA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zernusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glennda
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Heriman, John L
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376

```


; PRIOR FILING DATE: 2001-01-16
 ; NUMBER OF SEQ ID NOS: 269
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 25
 ; LENGTH: 1020
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-051-874-25

Query Match 20.5%; Score 225.6; DB 15; Length 1020;
 Best Local Similarity 58.0%; Pred. No. 1.3e-49;
 Matches 443; Conservative 0; Mismatches 309; Indels 12; Gaps 2;

QY	129	CTGCGGGCGCTGAGCCCTCGCCGCGATCGTGGGGGCTCAAAACGCGACGCCGGCAC	188
DB	171	CTGCGGGCGACCCCGCATGTCACGTGCGATGCGTGGGGCCGGGATGGCCGGACGGAGA	230
QY	189	CTGGCCCTTGGCAAGTGAAGCTGACCATGAGGTGGCCACATCTGGGGGGCTCCCTCAT	248
DB	231	GTGGCCGTGGCAGCGAGCATCAGCATCTGGGGCACAGTGTGGGGGGTCTGCTCAT	290
QY	249	CGCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTCATGAAGATGGAGCGTGGAGCC	308
DB	291	CGCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTCATGAAGATGGAGCGTGGAGCC	341
QY	309	CGCGCGGAGTGTGCGTACTGCTGGGCTGCACTCCAGGAGCGGCCCTTGGACGGCGC	368
DB	342	GCAGCTGAGTACCGGCTGCGCTGGGGGGCGCTGCTGTTGGGTCCACCTCGCCCCGAC	401
QY	369	GCACACCGCGAGTGGCCCTCATCTGTGTGCGGGCCAACTACAGCCAAAGTGGAGCTGGG	428
DB	402	GCTCTGGTCCCGTGGCAGGGTGTGCTGCTGCCCGGCACTACTCCGAGGACGGGCCCG	461
QY	429	CGCGGACCTTGGCCCTGTGCGCTTGGCTTCCCGGAGCGCTGGGCCCTGGCCCTGTGGCC	488
DB	462	CGGCGACCTGGCATCTGTCAGCTGCGTCCGCGCGGTGCCCTTGGAGGCTGCGCTCCAAAC	521
QY	489	TGCTGTGCTGCGCGCGCTTACACCGCTTCTGTCAGCGGACCGCTTGTGGGCCACCGG	548
DB	522	CGTGTGCTGCGCGCGCTTACACCGCTTCTGTCAGCGGACCGCTTGTGGGCCACCGG	591
QY	549	CTGGGAGAGCTTCAGGAGGAGATCTCTGCTCTCTCCCTTGGGTGTACAGGAAGTGA	608
DB	582	CTGGGCGAGCTTCGGCCCGAGGAGTCCCTCCAGAGTGGCGACCGCTACAAGAGTAAG	641
QY	609	GCTAAGGCTGTGGCGAGGCGACCTGTGATGTCTCTACAGCAGCGCCGCTCCCTTCAA	668
DB	642	GGTGGCGCTGTGGACTTCGCGCACCTCGGACCGGCTCTACCGTGGGCGCGGACGTGCC	701
QY	669	CTTCACTCTCCAGATAT---TGCCAGGAGTGTGTGTGCTGTACCCAGAGGGCCGCGAG	725
DB	702	CCAGGCTGAGGCGATGTGTGCTGCTGGAGTCTGTGTGCGGCTACCCAGGCGCACAA	761
QY	726	GGACACCTGCGAGGCTGACTCTGGGGGGCCCTTGGTCTGTGAGGAAGCGCGCTGGTT	785
DB	762	GGACGCTGCGAGGCTGATCTGGGGGACCTCTACCTGCTGCTGAGTCTGGGAGCTGGGT	821
QY	786	CCAGGAGGAATACACAGCTTGTGGTGTGGTGTGGAGGAGAAACCGCCCTGGAGTTT	845
DB	822	CTGTGGTGGGCTGTGAGTGTGGGCAAGGGTGTGTGCTTGCCTTGCCTTGCCTTGCCT	881
QY	846	CACCTGTGTGCTACCTATGAGGATGATACGGGAGCAGGTGA	889
DB	882	CACAGTGTGCGCACATATAGCCCTGGATTCAGGCTCGGTCA	925

RESULT 5
 US-10-041-400A-1
 ; Sequence 1, Application US/10041400A
 ; Publication No. US20020110895A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Darrow, Andrew
 ; APPLICANT: Andrade-Gordon, Patricia

; APPLICANT: Qi, Jenson
 ; TITLE OF INVENTION: DNA Encoding the Human Serine
 ; FILE OF INVENTION: Protease EOS
 ; FILE REFERENCE: ORT-1031
 ; CURRENT APPLICATION NUMBER: US/10/041,400A
 ; CURRENT FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US/09/387,375
 ; PRIOR FILING DATE: 1999-08-31
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1613
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-041-400A-1

Query Match 20.3%; Score 224; DB 13; Length 1613;
 Best Local Similarity 57.9%; Pred. No. 3.6e-49;
 Matches 442; Conservative 0; Mismatches 310; Indels 12; Gaps 2;

QY	129	CTGCGGGCGCTGAGCCCTCGGCCGCGATCGTGGGGGCTCAAAACGCGACGCCGGCAC	188
DB	149	CTGCGGGCGAGCCCCGCGATGTCAGTCCGATCGTGGGGCCGGGATGGCCGGAGCGAGA	208
QY	189	CTGGCCTTGGCAAGTGAAGCTGACCATGAGAGTGGCCACATCTGGGGGGCTCCCTCAT	248
DB	209	GTGGCCGTGGCAGGCGAGCATCCAGCATCTCGGGGCGACAGTGTGCGGGGGTCTGCTCAT	268
QY	249	CGCCCTCTCTGGGTCTCTCGGCTGCTCACTGTTTCATGAAGATGGAGCGTGGAGCC	308
DB	269	CGCCCTCTCTGGGTCTCTCGGCTGCTCACTGTTTCATGAAGATGGAGCGTGGAGCC	319
QY	309	CGCGGCGAGTGTGCGTACTGCTGGGCGTGCATCTCCAGGAGCGGCCCTTGGACGGCGC	368
DB	320	GCGAGTGAATACCGGTGCGCTGGGGCGCTGGCTGTGGGCTCCACCTCGCCCCGAC	379
QY	369	GCACACCGCGAGTGGCCCTCATCTGTGTGCGGGCCAACTACAGCCAAAGTGGAGTGGG	428
DB	380	GCTCTGCTGCGCTGCGAGCGGTGCTGCTGCTGCCCGGCACTCTCCAGGAGCGGGCCCG	439
QY	429	CGCGGACCTTGGCCCTGTGCGCTTGGCTTACCGCGCAGCTTGGGCCCGCCCGCTGGCC	488
DB	440	CGCGGACCTTGGCACTGTCTGCTGCTGCGCGGCGCTTGGCGCGCCCTTGGAGCGCTCGCGTCCAAAC	499
QY	489	TGCTGTGCTGCGCGCGCTTACACCGCTTCTGTGACAGCGCACCGCTTGTGGGCCACCGG	548
DB	500	CGTCTGCTGCGCGCGCTTACACCGCTTCTGTGACAGCGCACCGCTTGTGGGCCACCGG	559
QY	549	CTGGGAGAGCTTCAGGAGGCGAGTCTCTGCTGCTCTCCCTTGGGTGTCTACAGGAAGTGA	608
DB	560	CTGGGCGAGCTTCGCGCCAGAGTGTGCCCTTCCAGAGTGGCGGACCGCTTACAAGAGTAAG	619
QY	609	GCTAAGGCTGTGGGCGAGGCGACCTGTCAATGTCTCTACAGCAGCGCCGCTCCCTTCAA	668
DB	620	GGTGGCGCTGTGAGTCTGCGGACCTTGGCGACCGGCTTCTACAGTGGGCGCGGACGTGCC	679
QY	669	CTTCACTCTCCAGATAT---TGCCAGGAGTGTGTGTGCTGTGCTTACCCAGAGGGCCGCGAG	725
DB	680	CCAGGCTGAGCGCATTTGTGCTGCTGGAGTCTGTGTGCGGCTACCCCGAGGCGCACAA	739
QY	726	GGACACCTGCGAGGCTGACTCTGGGGGGCCCTTGGTCTGTGAGGAAGCGCGCTGGTT	785
DB	740	GGACGCTTGCAGGCTGATTTCTGGGGGAGCTCTGACCTGCTGCTGCTGCTGCTGCTGCT	799
QY	786	CCAGGAGGAATCACAGCTTGTGGTGTGGCTGTGACCGAGAAACCGCCCTGGAGTTT	845
DB	800	CTGTGGGCGGTGAGTGGGCGAGGGTGTGTGCTTGCCTTGCCTTGCCTTGCCTTGCCT	859
QY	846	CACCTGTGTGCTACCTATGAGGATGATACCGGAGCAGGTGA	889
DB	860	CACAGTGTGCGCACATATAGCCCTGGATTCAGGCTCGGTCA	903

RESULT 6

US-10-041-264A-1
; Sequence 1, Application US/10041264A
; Publication No. US20020142446A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,264A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-041-264A-1

Query Match 20.3%; Score 224; DB 13; Length 1613;
Best Local Similarity 57.9%; Pred. No. 3.6e-49;
Matches 442; Conservative 0; Mismatches 310; Indels 12; Gaps 2;
QY 129 CTGCGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAAACGCGACGCCGGGCAC 188
DB 149 CTGCGGGCGACCCCGCATGCTCAGTCGGATCGTGGGGCGCGGATGGCGGACGGAGA 208
QY 189 CTGGCCCTTGGCAAGTAGAGCTGCACCATGAGGTGGCCACATCTGGGGGGCTCCCTCAT 248
DB 209 GTGGCCGTGGCAGCGAGCATCAGCATCTTGGGGCACAGTGTGGGGGGTGGCTCAT 268
QY 249 CGCCCCCTCTCGGCTCTCTCCGCTGCTCATCTGTTTCATGACAAATGGAGCGTTGGAGCC 308
DB 269 CGCCCCCGAGTGGTGTGACACGCGCGCATCTCTTCCCGAGG-----GGGCACT 319
QY 309 CGCGCGCGAGTGGTGGTGTGCTGCTGGCGGTGCTACACGACCAAGTAGAGTGGG 428
DB 380 GCTCTCGGTCCCGTGGCAGCGGTGTGCTGCTGCCCGCGGACTCTCCGAGGACGGGCGCG 439
QY 429 CGCGACCTTGGCCCTGTGGCTTGCCTTCCCGCGAGCTTCCCGCGAGCTGGGCGCCGCTGTGGCC 488
DB 440 CGCGCGACCTGGCACTGTGTCAGTGTGCTGCGCGGTGCGCCCTGAGCGCTCGCGTCCAACC 499
QY 489 TGTCTGCTGCGCGCGCTTCAACCGCTTGTGTCAGCGCACCGCTGCTGGGCGCCACCGG 548
DB 500 CGTCTGCTGCGCGCGCTGCG 559
QY 549 CTGGGAGAGCTCCAGAGGACAGATCTCTGCTGCTCTCCCTGGGTGTCTACAGGAAGTGA 608
DB 560 CTGGGCGAGCTCCGCGCGAGGTGCGCTTCCAGAGTGGCGACCGCTACAAGAGTAAG 619
QY 609 GCTAAGGCTGTGGGCGAGCGCACCTGTCTCAATGTCTTACAGCAGCGCGCGTCCCTTCAA 668
DB 620 GGTGCGCGCTGTGAGCTCGCGCACCTGTGCGAGCGCGCTCTACCAAGTGGCGCGGACGTGCC 679
QY 669 CCTCACTTCCAGATAT---TGCCAGGGATGTGTGTGCTGCTACCCAGAGGGCGCGAG 725
DB 680 CCAGGCTGAGCGCATGTGTGCTGGGAGTGTGTGCGCGGTACCCCGCGGCGCCACAA 739
QY 726 GGACACCTGCCAGGGTACTCTGGGGGGCCCTGGTCTGTGAGAGAGCGCGCGCTGGTT 785
DB 740 GGACCGCTGCAGGGTGAATCTGGGGGACCTCTGACCTGCTGCTGCTGCTGGGAGTGGGT 799
QY 786 CCAGGCGAGGAATACCAAGCTTTGGGTTTGGCTGTGAGCGGAGAAACCGCGCTGGAGTTTT 845

DB 800 CTTGTGGGCGTGTGTGAGCTGGGGCAAGGGTGTGTCCCTGCCCAACCGTCCAGGGGTCTA 859
QY 846 CACTGCTGTGCTACTATGAGGATGATAGGAGTGGAGCAGGTGA 889
DB 860 CACCAGTGTGGCCACATATAGCCCCCTGGATTTCAGGCTGCGGTCA 903

RESULT 7

US-10-042-091A-1
; Sequence 1, Application US/10042091A
; Publication No. US20020142447A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/042,091A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-091A-1

Query Match 20.3%; Score 224; DB 13; Length 1613;
Best Local Similarity 57.9%; Pred. No. 3.6e-49;
Matches 442; Conservative 0; Mismatches 310; Indels 12; Gaps 2;
QY 129 CTGCGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGCTCAAACGCGACGCCGGGCAC 188
DB 149 CTGCGGGCGACCCCGCATGCTCAGTCGGATCGTGGGGCGCGGATGGCGGACGGAGA 208
QY 189 CTGGCCCTTGGCAAGTAGAGCTGCACCATGAGGTGGCCACATCTGGGGGGCTCCCTCAT 248
DB 209 GTGGCCGTGGCAGCGAGCATCAGCATCTTGGGGCACAGTGTGGGGGGTGGCTCAT 268
QY 249 CGCCCCCTCTCGGCTCTCTCCGCTGCTCATCTGTTTCATGACAAATGGAGCGTTGGAGCC 308
DB 269 CGCCCCCGAGTGGTGTGACACGCGCGCATCTCTTCCCGAGG-----GGGCACT 319
QY 309 CGCGCGCGAGTGGTGGTGTGCTGCTGGCGGTGCTACACGACGCGGCTTCCAGCGCGCG 368
DB 320 GCGAGTGTGATACCGGTGCGCTGCGGGGCGTGTGGTCTGGGCTCCACTCGCGCGCGAC 379
QY 369 GCACACCGCGCAGTGGCGCGCATCTGTGTGCGCGGCAACTACAGCAAGTGGAGTGGG 428
DB 380 GCTCTCGGTCCCGTGGCAGCGGTGTGCTGCTGCCCGCGGACTCTCCGAGGACGGGCGCG 439
QY 429 CGCGACCTTGGCCCTGTGGCTTGCCTTCCCGCGAGCTTCCCGCGAGCTGGGCGCCGCTGTGGCC 488
DB 440 CGCGCGACCTGGCACTGTGTCAGTGTGCTGCGCGGTGCGCCCTGAGCGCTCGCGTCCAACC 499
QY 489 TGTCTGCTGCGCGCGCTTCAACCGCTTGTGTCAGCGCACCGCTGCTGGGCGCCACCGG 548
DB 500 CGTCTGCTGCGCGCGCTGCG 559
QY 549 CTGGGAGAGCTCCAGAGGACAGATCTCTGCTGCTCTCCCTGGGTGTCTACAGGAAGTGA 608
DB 560 CTGGGCGAGCTCCGCGCGAGGTGCGCTTCCAGAGTGGCGACCGCTACAAGAGTAAG 619
QY 609 GCTAAGGCTGTGGGCGAGCGCACCTGTCTCAATGTCTTACAGCAGCGCGCGTCCCTTCAA 668
DB 620 GGTGCGCGCTGTGAGCTCGCGCACCTGTGCGAGCGCGCTCTACCAAGTGGCGCGGACGTGCC 679
QY 669 CCTCACTTCCAGATAT---TGCCAGGGATGTGTGTGCTGCTACCCAGAGGGCGCGAG 725
DB 680 CCAGGCTGAGCGCATGTGTGCTGGGAGTGTGTGCGCGGTACCCCGCGGCGCCACAA 739

Db 331 GCTCCCTGCGGTGTG-----GCCCCCAAGCAGCATCACAGTGGCAGAGTGCAGTC 384
Qy 181 CCGGGACACCTGGCTGGCAAGTAGAGCTGCACATGAGAGTGGCCACATCTCTCGGGGGC 240
Db 385 GCGGTCAGTGGCCCTGGCAGTGCAGATCACCTATGAAGCGTCCATGTGTGTGGTGGC 444
Qy 241 TCCCTCATCGCCCTCCCTGGGTCCTCTCGGTCCTCAGCTGCTCATGTTTCATGACGAATGGAGC 300
Db 445 TCTCTGTGTCTGAGCAGTGGGTGCTGTGCTGCTGCTCATGCTGCTTCCCGCGA----- 497
Qy 301 TTGGAGCCCGCGCCGAGTGGTCTGGTACTCTCTGGGCGTGCATCTCCAGGACGGGCCCTG 360
Db 498 --GCACACAAAGGAGGCTATGAGGTCAAGCTGGGGGCCACCCAGCTAGACTCTCTACTCC 555
Qy 361 GAGGGCGGCACACCCCGCAGTGGCGGCCATCGTGTGGCGGCCACTACACGCAAGT 420
Db 556 GAGGAGCCCAAGGTCAGCACCTTAAGGACATCATCCCCACCCAGCTACCTCCAGGAG 615
Qy 421 GAGTGGCGCCGACCTGGCCCTGCTGGCTGAGCTCAGCCGCGCAGCTGGGCCCGCC 480
Db 616 GGTCCCGAGGGGACATTGACTCTTCCAACTCAGCAGACCCATCAGCTTCTCCCGTAC 675
Qy 481 GTGTGGCTGTCTGCTGCGCCCGCGCTCACCGCTTCGTGCAAGGACCGGCTGTGG 540
Db 676 ATCGGGCCATCTGCTCCCTGCGAGCAACGCTCTTCCCAAGGCGCTCCACTGCACT 735
Qy 541 GCACCGGCTGGGAGAGCTCCAGGAGCAGATCTCTGCTCTCCCTCGGTGGTGTACAG 600
Db 736 GTCACTGGCTGGGTGCTATGTGGCCCCCTCAGTGGAGCTCTGAGCGCCCAAGCCACTG 795
Qy 601 GAAGTGAGCTAAGGCTGCTGGCGAGGCCACCTGTCAATGTCTCTACAGCAGCCCGCT 660
Db 796 CAACTCGAGGTGCTGTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGGCTTGT 855
Qy 661 CCCTTCAACCTCACTCTCCAGATATTGCCAGGGAGTGTGTGTGCTGCTACCCAGAGGC 840
Db 976 TGGTACTGAGCGGCAATGTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGGCTTGT 1035
Qy 841 GTTTTCACTGTGTGGTACTCTATGAGGCATGTGATACGGGAGCAGGTGTATGGGTTCAG 900
Db 1036 GTGTACACTCTGGCTCCAGCTATGCTCTCTGGATCCAAAGCAAGGTGACAGAACTCCAG 1095
Qy 901 CCTGGGCTGCTTTCCACCCAG 924
Db 1096 CCTGTGTGGTGCCCAACCCAG 1119

RESULT 10
US-09-967-768A-141
; Sequence 141, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIORITY APPLICATION NUMBER: US/60/236,109
; PRIORITY FILING DATE: 2000-09-28
; PRIORITY APPLICATION NUMBER: US/60/236,034
; PRIORITY FILING DATE: 2000-09-28
; PRIORITY APPLICATION NUMBER: US/60/236,111
; PRIORITY FILING DATE: 2000-09-28

; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 141
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-141
Query Match 20.1%; Score 221.2; DB 9; Length 1834;
Best Local Similarity 54.2%; Pred. No. 2e-48;
Matches 501; Conservative 0; Mismatches 408; Indels 15; Gaps 2;
Qy 1 GGGCCCTTCTCTGGGCCATGCCCCAGAAAGGGGTCTCTGGGGCTTGGGAGCTGGGGGCT 60
Db 211 GGGGCCCTTGTCTGGCCATGCCCCAGAAAGGGGTCTCTGGGGCTTGGGAGCTGGGGGCT 270
Qy 61 GTGGCCAAATCTGACTCATACTCAGCTTTAGGGTGTGGTCCGTCCTCCGACCCGCTAGGGGC 120
Db 271 GTGGCCATTTCTGCTCTATCTTGGATTACTCCGGTCTGGGACAGAGCGGGAAGGGGCGAA 330
Qy 121 CCCCCGTACTGCGGGCGCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAAAACGGCGCAG 180
Db 331 GCTCCCTGCGGTGTG-----GCCCCCAAGCAGCATCACAGTGGCAGCTGCGATC 384
Qy 181 CCGGGACCTGGCCCTTGGCAAGTAGAGCTGCACATGAGAGTGGGCCACATCTCGGGGGC 240
Db 385 GCGGTCAGTGGGCCCTGGCAGGTGAGCATCACCTATGAAGCGCTCCATGTGTGTGGTGGC 444
Qy 241 TCCCTCATCGCCCTCTCTGGGTGCTCTCGGCTGCTCAGCTGCTTTCATGAGCAATGGAGC 300
Db 445 TCTCTGTGTCTGAGCAGTGGGTGCTGCTGAGCTGCTCAGCTGCTTCCCGCGA----- 497
Qy 301 TTGGAGCCCGCGCCGAGTGGTCTGGTACTCTCTGGGCGTGCATCTCCAGGACGGGCCCTCTG 360
Db 498 --GCACCAAAAGGAGGCTATGAGGTCAAGCTGGGGGCCACCCAGCTAGACTCTCTACTCC 555
Qy 361 GAGCGCGGCACACCCCGCAGTGGCGCCATCGTGTGGTGGCGGCCAACTACAGCAAGT 420
Db 556 GAGGACCGCAAGGTGAGCACCCCTGAAGGACATCATCCCCCAGCCAGCTACCTCCAGGAG 615
Qy 421 GAGTGGCGCCGACCTGGCCCTGCGCTGCGCTGAGCTCACCCCGCAGCTGGGCCCGCC 480
Db 616 GGTCCCGAGGGGACATTTGCACTCTCCAACTCAGCAGACCCATCAGCTTCTCCCGTAC 675
Qy 481 GTGTGGCTGTGTGCTGCGCCCGCGCTCACACGCTTCGTGCAAGGACCGGCTGTCTGG 540
Db 676 ATCGGGCCATCTGCTCCCTGCGAGCAACGCTCTTCCCAAGGCGCTCCACTGCACT 735
Qy 541 GCACCGGCTGGGAGAGCTCCAGGAGCAGATCTCTGCTCTCCCTCGGTGGTGTACAG 600
Db 736 GTCACTGGCTGGGTGCTATGTGGCCCCCTCAGTGGAGCTCTGAGCGCCCAAGCCACTG 795
Qy 601 GAAGTGAGCTAAGGCTGCTGGCGAGGCCACCTGTCAATGTCTCTACAGCAGCCCGCT 660
Db 796 CAACTCGAGGTGCTGTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGGCTTGT 855
Qy 661 CCCTTCAACCTCACTCTCCAGATATTGCCAGGGAGTGTGTGTGCTGCTACCCAGAGGC 720
Db 856 AAGCCTGAGGAGCGCACTTTTCCAAAGAGACATGCTGTGTGTGCTGTGCTATGTGAGGGG 915
Qy 721 CCGAGGACACCTGCGAGGTGACTCTGGGGGGCCCTGTGCTGTGAGGAGGGCGCGC 780
Db 916 GCGAAGAGCGCTGCCAGGGTGACTCTGGGGGCCCACTCTCTCCCTGTGGAGGGTCTC 975
Qy 781 TGGTTCCAGGAGGAATCACAGCTTTTGGGTGTGGCTGTGGAGCGGAAACCGCCCTGGA 840
Db 976 TGGTACTGAGCGGCAATGTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGGCTTGT 1035
Qy 841 GTTTTCACTGTGTGGTACTCTATGAGGCATGTGATACGGGAGCAGGTGTATGGGTTCAG 900
Db 1036 GTGTACACTCTGGCTCCAGCTATGCTCTCTGGATCCAAAGCAAGGTGACAGAACTCCAG 1095
Qy 901 CCTGGGCTGCTTTCCACCCAG 924

LOCATION: (1565)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (1598)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (1620)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-208

Query Match 20.0%; Score 220.6; DB 9; Length 1668;
Best Local Similarity 54.7%; Pred. No. 2.8e-48;
Matches 506; Conservative 2; Mismatches 401; Indels 16; Gaps 3;
QY 1 GGGCCCTTCTCTGGCCATGCGCCAGAGGGGTCTCTGGGCTTGGGCGAGCTGGGGCT 60
DB 191 GGGCCCTTCTCTGGCCATGCGCCAGAGGGGTCTCTGGGCTTGGGCGAGCTGGGGCT 250
QY 61 GTGGCCAAATCTGACTCATCTACTTTACGGGTGTGTCCCTCGGACCCCTAGGGGC 120
DB 251 GTGGCCAAATCTGACTCATCTACTTTACGGGTGTGTCCCTCGGACCCCTAGGGGC 310
QY 121 CCCCCGTACTGGGGCCCTGAGCCCTCGGCCGATCTGGGGGCTCAACAGGCGAG 180
DB 311 GTCYCTCGGGTGTG-----GCCCCCAAGCAGCATCACAGGTGGCAGCTGCAGTC 364
QY 181 CCGGCACCTGGCCCTGGCAAGTAGCCTGCACATGAGGTGGCCACATCTGGGGGC 240
DB 365 GCGGTCACTGGCCCTGGGAGTGCAGCATCACTATGAAGGCTCATGTGTGTGGC 424
QY 241 TCCCTCATCGCCCTCTCTGGGTCTCTCGGTGTCTACTGTTCATGACGAATGGGACG 300
DB 425 TCTCTGTGTCTGAGCAGTGGGTGTGTCTGAGTGTCTCACTGTCTCCCGCAGCA----- 477
QY 301 TTGGAGCCCGCCGAGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 478 --GCACCAAGAAAGCCATGAGGTCAAGTGTGGGGCCCAAGTACACTCTCTACTCC 535
QY 361 GACGGCGCCACACCCCGCAGTGGCGGCATCGTGGTGGCGGCATCGTGGTGGCGGC 420
DB 536 GAGGAGCCCAAGGTGAGCACCTTGAAGGACATCATCCCCAACCAGCTACCTCCAGGAG 595
QY 421 GAGTGGGCGCCGACCTGGCCCTCTGCGCTGGCCCTCAACCGCCAGCTGGGCCCGCC 480
DB 596 GGTCTCCAGGGCGACATGCACTCTCTCCAACTCAGCAGACCCATCACTCTCTCCCGCTAC 655
QY 481 GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
DB 656 ATCCGGCCCATCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 715
QY 541 GCCACCGGTGGGGAGACGTCCAGGAGGAGATCCCTCTGCTCTCTCTCTCTCTCTCTCT 600
DB 716 GTCACTGGCTGGGGTCACTGTGGCCCTCTAGTGAGCTCTCTGAGCCCAAGCACTGGAG 775
QY 601 GAAGTGGAGCTAAGGTCTGTGGCGAGGCCACTGTCAA--TGTCTTACAGCAGCCCGG 659
DB 776 CAACCTGAGGTGCTCTGATCATGCTGAGACGTGGTAACTGCTGTACAACTGACGCG 835
QY 660 TCCCTTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGTGTGGTATCCAGAGGG 719
DB 836 CAAGCTGAGGAGCCGACCTTTGTCCAAGAGGACATGGTGTGTGTGGTATGTGAGGG 895
QY 720 CCGCAGGAGCACCTGCGAGGGTCACTCTGGGGGGCCCTGGTCTGTGAGGAAGCGGCGG 779
DB 896 GGGCAGGAGCCCTGCCAGGGTCACTCTGGGGGCCCACTCTCTGCTGCTGTGTGAGGGTCT 955
QY 780 CTGGTTCAGGAGGAATACACAGCTTTGGTTTGGCTGTGAGCAGGAGAAACCGCCCTGG 839
DB 956 CTGCTACCTGACGGGCAATTGTGAGCTGGGAGATGCTGTGTGGGCGCCGCAACAGGCTGG 1015
QY 840 AGTTTTCACCTGTGTGGCTACCTATGAGGATGATGAGGAGCAGGTGATGGGTTTCA 899
DB 1016 TGTGTACACTCTGGCTTCCAGCTATGCTCTCTGATCCAAAGCAAGGTGACAGAACTCCA 1075

QY 900 GCCTGGGCTGCTTCCATCCACCCAG 924
DB 1076 GCCTGCTGGTGGCCCAACCCAG 1100

RESULT 13

US-10-041-400A-8
Sequence 8, Application US/10041400A
Publication No. US20020110895A1
GENERAL INFORMATION:
APPLICANT: Barrow, Andrew
APPLICANT: Andrade-Gordon, Patricia
APPLICANT: Qi, Jenson
TITLE OF INVENTION: DNA Encoding the Human Serine
TITLE OF INVENTION: Protease EOS
FILE REFERENCE: ORT-1031
CURRENT APPLICATION NUMBER: US/10/041,400A
CURRENT FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US/09/387,375
PRIOR FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 1130
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid
OTHER INFORMATION: sequence of EOS zymogen fusion gene
US-10-041-400A-8

Query Match 19.7%; Score 216.8; DB 13; Length 1130;
Best Local Similarity 58.0%; Pred. No. 2.7e-47;
Matches 427; Conservative 0; Mismatches 297; Indels 12; Gaps 2;
QY 157 ATCTGGGGGGCTCAACGCGCAGCGGGGCACCTGCGCTTGGCAAGTAGCTGACCAT 216
DB 166 ATCTGGGGGGCTATGCTCTAGAGGACGAGAGTGGCGTGGCAGGCGAGCATCCAGCAT 225
QY 217 GAGGTGGGCAATCTGGGGGGCTCTCTCATCGCCCTCTCTCTCTCTCTCTCTCTCTCT 276
DB 226 CCGTGGGCGACAGTGTGGGGGGTCTCTCATCGCCCTCTCTCTCTCTCTCTCTCTCTCT 285
QY 277 CACTGTTTCAACCAATGGGACGTTGGAGCCGCGCGAGTGGTGGTACCTCTCTCTCTCTCT 336
DB 286 CACTGTTTCCCGAGCA-----GGGCACTGCGAGTGAATACCGGTGGCTGGGG 336
QY 337 GTGCACTCCAGGACGGGCCCTGTGACCGCGCGCACACCCGCGCAGTGGCCGCCATCGTG 396
DB 337 GCGCTGCGTCTGGGCTCCACCTCGCCCGCGCAGCTCTCGGTGCGCGTGCAGCGGTGCTG 396
QY 397 GTGCGGCGCACTACAGCAAGTGGAGTGGGGCGGACCTGGCCCTCTCTCTCTCTCTCTCT 456
DB 397 CTGCGCCCGGAGCTACTCGAGGACCGGGCCCGCGGACCTGGCACCTGCTGCTGAGTGGT 456
QY 457 TCACCGCGCAGCTGGGGCCCGCGCTGTGGCTGTCTGCTCTCTCTCTCTCTCTCTCTCTCT 516
DB 457 CCGCGGCTGCGCTCTGAGCGCTCGCTCCAAACCGCTCTGCTGCTGCTGCTGCTGCTGCTG 516
QY 517 TTCTGTGACGGCACCGCTCTGTGGGCCACCGGCTGGGAGAGCTCCAGGAGGACATCTCT 576
DB 517 CCGCGCGCGGCGACACCATGCGGGTCAACCGGCTGGGGCAGGCTCCGCGCCAGAGTGCC 576
QY 577 CTGCTCTCTCTCTGGTGTACAGGAAGTGGAGCTAAGGCTGTGGCGGAGCCACCTCTCT 636
DB 577 CTCTCAGAGTGGCGACCTCTACAGGAGTAAGGGTGGCGCTGCTGCTGCTGCTGCTGCTG 636
QY 637 CAATGTCTCTACAGCAGCGCGCTCTCTCAACCTCACTCTCTCAAGATAT---TGCCAGGG 693
DB 637 GACGCGCTCTACCACTGGGCGCGAGCTGTGCCCGAGGTGAGCGCATTTGTGTCTGCTGG 696
QY 694 ATGCTGTGTGCTGCTATCCAGAGGGCGCGAGGAGCACTGTCCAGGGTGACTCTCTGGGGG 753


```

Db 697 AGTCTGTGTGCGGCTACCCCGAGGCGCAAGAGACGCTTGCAGGGTGATTTCTGGGGGA 756
QY 754 CCCTGGTCTGTAGGAGCGCGCTGTCTTCCAGCAGGATACACAGCTTTTGGGTTT 813
Db 757 CCTCTGACCTGCTGCAGTCTGGAGCTGGGTCTCTGGTGGCGTGTGTGAGCTGGGGCAAG 816
QY 814 GGCTGTGGACGAGAAACCCCTGGAGTTTTCACCTGTCTGTGCTTACCTATAGGCGATGG 873
Db 817 GGTGTGCTTGTCCCAACCTTCCAGGGTCTACACAGTGTGGCCACATATAGCCCTGG 876
QY 874 ATACGGGAGCAGTGA 889
Db 877 ATTACGGCTCGCGTCA 892

RESULT 14
US-10-041-264A-8
; Sequence 8, Application US/10041264A
; Publication No. US2002014246A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,264A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1130
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid
; OTHER INFORMATION: sequence of EOS zymogen fusion gene
US-10-041-264A-8
Query Match 19.7%; Score 216.8; DB 13; Length 1130;
Best Local Similarity 58.0%; Pred. No. 2,7e-47;
Matches 427; Conservative 0; Mismatches 297; Indels 12; Gaps 2;

QY 157 ATCGTGGGGGCTCAACAGCGCAGCCGCGACCTTGGGCTGGGAGTGGAGCTGCACCAT 216
Db 166 ATCGTTGGGGGCTATGCTCTAGAGGAGGAGAGTGGCGGTGGCAGCGAGCATCCAGCAT 225
QY 217 GGAGGTGGCCACATCTGTGGGGGGCTCCCTCATATGCCCGCTTCTGGGTCTCTCGCTGCT 276
Db 226 CCTGGGGGCAACGCTGTGGGGGGTCTGCTCATGCCCGCTTCTGGGTCTCTGACAGCGGG 336
QY 337 GTGCACTTCCAGGAGGAGTGGAGCGGCGCGCACACCGCGAGTGGTCTGCTGGGCTGCTG 396
Db 337 GCCTGTGCTGGGCTTCCACCTGCGCGCGCACCTGCGCGCGAGTGGTCTGCTGGGCTGCTG 396
QY 397 GTGCGGCGCAACTACAGAGTGGAGCTGGGCGCGCGAGCTGGGCTGCTGGGCTGCTGGGCT 456
Db 397 CTGCGGCGGCTACTCCGAGGAGCGGGGCGCGCGAGCTGGGCTGCTGGGCTGCTGGGCTG 456
QY 457 TCAACCGGCGAGCTGGGCGCGCGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 516
Db 457 CGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 516
QY 517 TTGCTGACGCGACCGCTGTGCGGCGCGCGGCTGCGGCGAGCTGCGAGGAGGAGTCTT 576
Db 517 CCGCCGCGCGGCAACCATGCGGCTGCGGCTGCGGCGAGCTGCGGCGAGCTGCGGCTGCGG 576

```

```

QY 577 CTGCTCTCCCTGGTGTCTACAGGAGTGGAGCTAAAGCTGCTGGCGAGGCGACCTCT 636
Db 577 CTCCAGAGTGGGAGCCGCTACAGGAGTAAAGGTGCGGCTGCTGAGACTCGCGCACCTGC 636
QY 637 CAATGTCTCTACAGCAGCCCGTCCCTTCAACCTCACTCTCCAGATAT---TGCCAGGG 693
Db 637 GACGGCTCTACACAGCTGGGCGCGAGCTGCCAGGCTGAGCGCATTTGTGCTGCCCTGGG 696
QY 694 ATGCTGTGTGCTGCTTACCCAGAGGCGCGCAGGAGACCTGCCAGGCTGACTCTGGGGGG 753
Db 697 AGTCTGTGTGCGGCTTACCCAGGCGCCACAGGAGCGCTGCCAGGGTGAATCTTGGGGGA 756
QY 754 CCCCTGCTGTGAGGAGGCGCGCTGCTTCCAGGCGAGGAATCACACAGCTTTTGGGTTT 813
Db 757 CCTCTGACCTGCTGCAGTCTGGGAGCTGGGTCTCTGGTGGCGTGGTGGGCAAG 816
QY 814 GGCTGTGGACGAGAAACCGCCCTGGAGTTTTCACCTGTGTGGCTACCTATAGAGGATGG 873
Db 817 GGTGTGCTTGTGCCCAACCGTCCAGGGGTCTACACAGTGTGGCCACATATAGCCCTGG 876
QY 874 ATACGGGAGCAGTGA 889
Db 877 ATTACGGCTCGCGTCA 892

RESULT 15
US-10-042-091A-8
; Sequence 8, Application US/10042091A
; Publication No. US2002014247A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/042,091A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1130
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid
; OTHER INFORMATION: sequence of EOS zymogen fusion gene
US-10-042-091A-8
Query Match 19.7%; Score 216.8; DB 13; Length 1130;
Best Local Similarity 58.0%; Pred. No. 2,7e-47;
Matches 427; Conservative 0; Mismatches 297; Indels 12; Gaps 2;

QY 157 ATCGTGGGGGCTCAAAACGCGCAGCCGCGACCTTGGCCTTGGCAAGTGGAGCTGCACCAT 216
Db 166 ATCGTTGGGGGCTATGCTCTAGAGGAGGAGAGTGGCGGTGGCAGCGAGCATCCAGCAT 225
QY 217 GGAGGTGGCCACATCTGTGGGGGGCTCCCTCATATGCCCGCTTCTGGGTCTCTCGCTGCT 276
Db 226 CCTGGGGGCAACGCTGTGGGGGGTCTGCTCATGCCCGCTTCTGGGTCTCTGACAGCGGG 285
QY 277 CACTGTTTTCATGAGAAATGGAGCGTGGAGCGCGCGCGAGTGGTCTGCTGCTGGGCT 336
Db 286 CACTGCTTCCCGAGGA-----GGGCACTGCCAGCTGAGTACCGCTGCGGCTGGGG 336
QY 337 GTGCACTTCCAGGAGCGGGGCTTGGAGCGGCGCGCACACCGCGAGTGGGCGCGCATCTGTG 396
Db 337 GCCTGTGCTGGGCTTCCACCTGCGCGCGCACCTGCGCGCGAGTGGTCTGCTGGGCTGCTG 396
QY 397 GTGCCGCGCAACTACAGAGTGGAGCTGGGCGCGCGAGCTGGGCTGCTGGGCTGCTGGGCT 456

```


Db	397	CTGCCCCCGGACTACTCCGAGGACGGGGCCCGCGCGACCTTGGCACTGCTGCAGCTGGT	456
Qy	457	TCACCCGCGCAGCCTGGGCGCGCGTGTGGCGCTGTCTGCTGCGCCCGCGCTCACACCGC	516
Db	457	CGCCCGGTGCCCCCTGAGGGCTCGCGGTCCAAACCGCTCTGCTTGGCCCGCGCGCGC	516
Qy	517	TTCTGTGACCGGCACCGCGCTGTGGGCGCACCGCGCTGGGAGACGTCCAGGAGGCAGATCCT	576
Db	517	CCGCGCGCGCGCACACCATGCCGGTCAACCGCTGGGGCAGCCTCCGCGCAGGAGTGCCC	576
Qy	577	CTGCGCTCTCCCGCTGAGGAGTGAAGCTAAGGCTGCTGGGCGAGGCCACCTGT	636
Db	577	CTCCAGAGTGGCGACCGCTACAAGGAGTAAGGGTGGCGCTGCTGGACTCCGCGCACCTGC	636
Qy	637	CAATGCTCTACAGCAGCGCGGTCCCTTCAACCTCACTCTCCAGATAT---	693
Db	637	GACGGCCTCTACAGTGGGCGCGGACGTGCCCGCAGGCTGAGCGCATGTGCTGCTGGG	696
Qy	694	ATGCTGTGTGTGGCTACCCAGAGGCGCGCAGGACACCTGCCAGGGTGA	753
Db	697	AGTCTGTGTGCGGCTACCCCGCGCCACAGGACGCTGCCAGGGTGA	756
Qy	754	CCCCTGCTGTGAGGAAGGCGCGCTGGTTCCAGGCAGGAATCACAGCTTTGGGTTT	813
Db	757	CCTCTGACCTGCTGCACTGTGGGAGCTGGGTCCCTGTGGGCGTGGTGGGCGCAAG	816
Qy	814	GGCTGTGAGCGAGAAACCGCGCTGGAGTTTTCATGCTGTGGCTACCTATGAGGCATGG	873
Db	817	GGTGTGCCCCCGCCACCGTCCAGGGTCTACACCACTGTGGCCACATATAGCCCCCTGG	876
Qy	874	ATACGGGAGCAGGTGA	889
Db	877	ATCAGGCTCGGTCA	892

Search completed: February 25, 2004, 12:28:25
Job time : 457 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 11:19:55 ; Search time 4327 seconds
(without alignments)
11038.600 Million cell updates/sec

Title: US-10-037-417-45
Perfect score: 1102
Sequence: 1 gggcccttctctgggcat.....ggggttctgaggggcctcc 1102

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.frod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1102	100.0	1102	6	AX512287	AX512287 Sequence
2	1102	100.0	1102	6	AX512289	AX512289 Sequence
3	716	65.0	2457	6	AX360096	AX360096 Sequence
4	716	65.0	2681	6	AX480935	AX480935 Sequence
5	631	57.3	2810	6	BD127529	BD127529 Primer fo
6	631	57.3	2810	9	AK075142	AK075142 Homo sapi
7	486	44.1	537	6	AX342934	AX342934 Sequence
8	384	34.8	456	6	AX342936	AX342936 Sequence
9	346	31.4	670	6	BD125219	BD125219 Primer fo
10	346	31.4	670	6	BD126485	BD126485 Primer fo
11	270	24.5	127769	9	AC009088	AC009088 Homo sapi
12	270	24.5	195476	2	AC135044	AC135044 Homo sapi
13	87	7.9	168064	9	AC093520	AC093520 Homo sapi
14	67	6.1	1796	6	AX098193	AX098193 Sequence
15	67	6.1	1800	9	BC001462	BC001462 Homo sapi
16	67	6.1	1835	6	AX098215	AX098215 Sequence
17	67	6.1	7008	9	HSU33446	U33446 Human prost
18	67	6.1	195476	2	AC135044	AC135044 Homo sapi
19	53	4.8	1726	6	AX675579	AX675579 Sequence
20	53	4.8	1834	6	AX335777	AX335777 Sequence
21	53	4.8	1834	6	AX336076	AX336076 Sequence
22	53	4.8	1834	6	AX336340	AX336340 Sequence
23	53	4.8	1834	6	AX409567	AX409567 Sequence
24	53	4.8	1834	6	AX474697	AX474697 Sequence
25	53	4.8	1834	9	HUMPROS	I41351 Homo sapien
26	51	4.6	596	6	AX193364	AX193364 Sequence
27	50	4.5	6207	9	BC035589	BC035589 Homo sapi
28	50	4.5	7604	9	AB002294	AB002294 Human mRN
29	50	4.5	182230	9	AC135050	AC135050 Homo sapi
30	49	4.4	1161	6	AX675581	AX675581 Sequence
31	35	3.2	64328	2	AC124816	AC124816 Mus muscu
32	35	3.2	71177	2	AC101465	AC101465 Mus muscu
33	35	3.2	185788	10	AC124461	AC124461 Mus muscu
34	35	3.2	219763	10	AC093175	AC093175 Mus muscu
35	35	3.2	232119	2	AC106629	AC106629 Rattus no
36	35	3.2	244161	2	AC117170	AC117170 Rattus no
37	26	2.4	26	6	AX512409	AX512409 Sequence
38	25	2.3	25	6	AX556574	AX556574 Sequence
39	25	2.3	28628	2	AC020159	AC020159 Drosophil
40	25	2.3	184272	3	AC009211	AC009211 Drosophil
41	25	2.3	221830	3	AC009393	AC009393 Drosophil
42	25	2.3	247955	3	AE003698	AE003698 Drosophil
43	24	2.2	402	6	AX369313	AX369313 Sequence
44	24	2.2	610	6	AR263844	AR263844 Sequence
45	24	2.2	995	4	DOGMCTRPA	M24664 Dog mast ce

ALIGNMENTS

RESULT 1	AX512287	Sequence 43 from Patent WO02053742.	1102 bp	DNA	linear	PAT 27-SEP-2002
AX512287	LOCUS	AX512287				
DEFINITION	AX512287	Sequence 43 from Patent WO02053742.				
ACCESSION	AX512287					
VERSION	AX512287.1	GI:23392631				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
1						
Kekuda,R., Alsobrook,J.P., Tchernev,V.T., Liu,X., Spyttek,X.A.,						
Patturajan,M., Grosse,W.M., Lepley,D.M., Burgess,C.E., Vernet,C.A.,						
Li,L., Gorman,L., Edinger,S., Scioire,P., Ellerman,K., Malyankar,U.,						

TITLE Rothenberg,M., Stone,D., Boldog,F., Shenoy,S. and Anderson,D.
 Proteins and nucleic acids encoding same
 Patent: WO 02053742-A 43 11-JUL-2002;
 Curagen Corporation (US)

FEATURES
 source 1..1102
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1102; DB 6; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTGTGCTGGGCGATGGCCAGAGGGGTCTGGGGCTGGGCGAGCTGGGGGCT 60
 DB 1 GGGCCCTGTGCTGGGCGATGGCCAGAGGGGTCTGGGGCTGGGCGAGCTGGGGGCT 60

QY 61 GTGGCCAAATCTGACTCATATCACTTACGGGTGGTGCCGCTCCGAGACCCGCTAGGGC 120
 DB 61 GTGGCCAAATCTGACTCATATCACTTACGGGTGGTGCCGCTCCGAGACCCGCTAGGGC 120

QY 121 CCCCCGTACTGGGGCGCCCTGAGCCCTCGGCCGCACTGTGGGGGGCTCAAAACGGCGAG 180
 DB 121 CCCCCGTACTGGGGCGCCCTGAGCCCTCGGCCGCACTGTGGGGGGCTCAAAACGGCGAG 180

QY 181 CCGGGCACTTGGCTTGGCAAGTGAGCTGCACCATGGAGGTGGCCACATCTCGGGGGC 240
 DB 181 CCGGGCACTTGGCTTGGCAAGTGAGCTGCACCATGGAGGTGGCCACATCTCGGGGGC 240

QY 241 TCCTCATTCGCCCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGT 300
 DB 241 TCCTCATTCGCCCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGT 300

QY 301 TTGGAGCCCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
 DB 301 TTGGAGCCCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360

QY 361 GAGCGCGCGCACACCCCGCAGTGGCGCCATCGTGGTGGCGCCCACTACAGCCCAAGTG 420
 DB 361 GAGCGCGCGCACACCCCGCAGTGGCGCCATCGTGGTGGCGCCCACTACAGCCCAAGTG 420

QY 421 GAGTGGGCGCGACACCTGGCGCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
 DB 421 GAGTGGGCGCGACACCTGGCGCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480

QY 481 GTGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
 DB 481 GTGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540

QY 541 GCCACCGGCTGGGAGACGTCCAGGAGGACAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
 DB 541 GCCACCGGCTGGGAGACGTCCAGGAGGACAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 600

QY 601 GAAGTGAAGTAAAGTGTGGGGAGGCCCACTGTCAATGTCTTACAGCCAGCCCGGT 660
 DB 601 GAAGTGAAGTAAAGTGTGGGGAGGCCCACTGTCAATGTCTTACAGCCAGCCCGGT 660

QY 661 CCTTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
 DB 661 CCTTCAACCTCACTCTCCAGATATTGCCAGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720

QY 721 CGCAGGACACCTGCCAGGGTGAATCTGGGGGGGCCCTGCTGTGTGAGGAAGCGGCGCG 780
 DB 721 CGCAGGACACCTGCCAGGGTGAATCTGGGGGGGCCCTGCTGTGTGAGGAAGCGGCGCG 780

QY 781 TGGTTCAGGAGGAATCAACAGCTTTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 840
 DB 781 TGGTTCAGGAGGAATCAACAGCTTTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 840

QY 841 GTTTTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900

DB 841 GTTTTCACTCTGTGCTACCTATGAGGCATGAGATACGGAGCAGGTGATGGTTCAGAG 900
 QY 901 CTTGGGCTTGCCTTTCCACCCAGCCCAAGAGACCCAGTCTGAGTTGTTTACATCAAAACG 960
 DB 901 CTTGGGCTTGCCTTTCCACCCAGCCCAAGAGACCCAGTCTGAGTTGTTTACATCAAAACG 960

QY 961 GCAATTCCTGGATTCTGCCAGAAATCCTTTTGGGCCCTTGTCCCATATATCAGTAGGAGTC 1020
 DB 961 GCAATTCCTGGATTCTGCCAGAAATCCTTTTGGGCCCTTGTCCCATATATCAGTAGGAGTC 1020

QY 1021 TCAATGGGACCAAAAGCCTTGTCTCCCTCGCTCTCTCCACACTCTCTCTCTGGGCGCTC 1080
 DB 1021 TCAATGGGACCAAAAGCCTTGTCTCCCTCGCTCTCTCCACACTCTCTCTCTGGGCGCTC 1080

QY 1081 TGGGGGTTCTGATGGGCGCTCC 1102
 DB 1081 TGGGGGTTCTGATGGGCGCTCC 1102

RESULT 2
 AX512289 1102 bp DNA linear PAT 27-SEP-2002
 LOCUS Sequence 45 from Patent WO02053742.
 DEFINITION AX512289
 ACCESSION AX512289
 VERSION AX512289.1 GI:23392632
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Kekuda,R., Alsbroook,J.F., Tchernev,V.T., Liu,X., Spytek,K.A.,
 Patturajan,M., Grosse,W.M., Lepley,D.M., Burgess,C.E., Vernet,C.A.,
 Li,L., Gorman,L., Edinger,S., Sciore,P., Ellerman,K., Malyankar,U.,
 Rotenberg,M., Stone,D., Boldog,F., Shenoy,S. and Anderson,D.
 TITLE Proteins and nucleic acids encoding same
 JOURNAL Patent: WO 02053742-A 43 11-JUL-2002;
 FEATURES Curagen Corporation (US)
 source 1..1102
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 100.0%; Score 1102; DB 6; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTCTCTCGGCGCATGGCCAGAGGGGTCTGGGGCTGGGCGAGCTGGGGGCT 60
 DB 1 GGGCCCTTCTCTCGGCGCATGGCCAGAGGGGTCTGGGGCTGGGCGAGCTGGGGGCT 60

QY 61 GTGGCCAAATCTGACTCATATCACTTACGGGTGGTGCCGCTCCGAGACCCGCTAGGGC 120
 DB 61 GTGGCCAAATCTGACTCATATCACTTACGGGTGGTGCCGCTCCGAGACCCGCTAGGGC 120

QY 121 CCCCCGTACTGGGGCGCCCTGAGCCCTCGGCCGCACTGTGGGGGGCTCAAAACGGCGAG 180
 DB 121 CCCCCGTACTGGGGCGCCCTGAGCCCTCGGCCGCACTGTGGGGGGCTCAAAACGGCGAG 180

QY 181 CCGGGCACTTGGCTTGGCAAGTGAGCTGCACCATGGAGGTGGCCACATCTCGGGGGC 240
 DB 181 CCGGGCACTTGGCTTGGCAAGTGAGCTGCACCATGGAGGTGGCCACATCTCGGGGGC 240

QY 241 TCCTCATTCGCCCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGT 300
 DB 241 TCCTCATTCGCCCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGTCTCTCGGT 300

QY 301 TTGGAGCCCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
 DB 301 TTGGAGCCCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360

462	QY	CGCACGCTGGGCCCGCCGCTGTGGCTGTCTGCTTGCCTGCCCCCGCCTCACACCCTTCGT	521
476	Db	CGCCAGCTTGGGCGCCGCGCTGTGGCTGTCTGCTTGCCTGCCCCCGCCTCACACCCTTCGT	535
522	QY	GCAACGACCCGCTGTCTGGCCACACGGCTGGGGAGACGATCCACAGGAGGACAGATCCCTCTGCC	581
536	Db	GCAACGACCCGCTGTCTGGCCACACGGCTGGGGAGACGATCCACAGGAGGACAGATCCCTCTGCC	595
582	QY	TCTCCCTCGGTTCTACAGGAAGTGGAGCTAAGGGCTGTGGSCGAGGCCACCTGTCAATG	641
596	Db	TCTCCCTCGGTTCTACAGGAAGTGGAGCTAAGGGCTGTGGSCGAGGCCACCTGTCAATG	655
642	QY	TCTCTACAGCAGCCCGGCTCCCTTCAACTCATCTCTCAGATATTGTCAGGGATGCTGTG	701
656	Db	TCTCTACAGCAGCCCGGCTCCCTTCAACTCATCTCTCAGATATTGTCAGGGATGCTGTG	715
702	QY	TGCTGGCTACCCAGAGGGCGCGAGGGAGACCTGCCAGGGTGACTCTGGGGGGCCCCTGGT	761
716	Db	TGCTGGCTACCCAGAGGGCGCGAGGGAGACCTGCCAGGGTGACTCTGGGGGGCCCCTGGT	775
762	QY	CTGTGAGGAAGCGCGCGCTGTGTTCCAGGCAGGAATCACACAGCTTTGGTTTTGGCTGTGG	821
776	Db	CTGTGAGGAAGCGCGCGCTGTGTTCCAGGCAGGAATCACACAGCTTTGGTTTTGGCTGTGG	835
822	QY	ACGAGAAACCGCCCTGGAGTTTCTACTGCTGTGGCTACTATAGGCAATGATACGGGA	881
836	Db	ACGAGAAACCGCCCTGGAGTTTCTACTGCTGTGGCTACTATAGGCAATGATACGGGA	895
882	QY	GCAGGTGATGGTTTCAGAGCCTGGGCGCTGCTTTCCACCCAGCCCGCCAGAACCCAGTC	941
896	Db	GCAGGTGATGGTTTCAGAGCCTGGGCGCTGCTTTCCACCCAGCCCGCCAGAACCCAGTC	955
942	QY	AGAT 945	
956	Db	AGAT 959	

RESULT 6	
AK075142	
LOCUS	2810 bp mRNA linear PRI 03-SEP-2002
DEFINITION	Homo sapiens cDNA FLJ30661 fis, clone PLACE1005003, weakly similar to PROSTASIN PRECURSOR (BC 3.4.21.-).
ACCESSION	AK075142
VERSION	GI:22761040
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Isogai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T., Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S., Kawai-Hio, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Kojima, S., Nagahara, K., Masuko, Y., Cho, I., Okano, K., Yoshikawa, Y., Aotsuka, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and Ninomiya, K.
TITLE	NEDO human cDNA sequencing project
JOURNAL	Unpublished
REFERENCE	2 (Bases 1 to 2810)
AUTHORS	Isogai, T. and Otsuki, T.
TITLE	Direct Submission
JOURNAL	Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3966)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure. Human Genome Center; cDNA 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).
FEATURES	Location/Qualifiers
SOURCE	1..2810

ORGANISM Homo sapiens
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 670)
 Ota.T., Nishikawa.T., Isogai.T., Hayashi.K., Ishii.S., Kawai.Y.,
 Wakamatsu.A., Sugiyama.T., Nagai.K., Kojima.S., Otsuki.T. and
 Koga.H.
 TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL HELIX RESEARCH INSTITUTE
 COMMENT OS Homo sapiens (human)
 PN JP 2002017375-A/650
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC
 C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
 10, C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source 1..670
 FT /organism='Homo sapiens (human)'.
 FEATURES
 source Location/Qualifiers
 1..670
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'
 ORIGIN
 Query Match 31.4%; Score 346; DB 6; Length 670;
 Best Local Similarity 99.7%; Pred. No. 6.3e-171;
 Matches 396; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 162 GGGGGGCTCAAAACGCGAGCGCGGACCTGGCCCTGGCAAGTGGAGCTGACCATGGAGG 221
 DB 176 GGGGGGCTCAAAACGCGAGCGCGGACCTGGCCCTGGCAAGTGGAGCTGACCATGGAGG 235
 QY 222 TGGCCACATCTGGGGGGCTCCCTCATCGCCCGCTCTGGGTCTCTCCGCTGCTCACTG 281
 DB 236 TGGCCACATCTGGGGGGCTCCCTCATCGCCCGCTCTGGGTCTCTCCGCTGCTCACTG 295
 QY 282 TTTTCATGCAATGGGAGCTTGGAGCGCGCGAGTGGTTCGCTACTGCTGGCGGTGCA 341
 DB 296 TTTTCATGCAATGGGAGCTTGGAGCGCGCGAGTGGTTCGCTACTGCTGGCGGTGCA 355
 QY 342 CTCCAGGACGGGGCCCTGGAGCGCGCGACACCCGCGCAGTGGCGGCTGCTGGTGGCC 401
 DB 356 CTCCAGGACGGGGCCCTGGAGCGCGCGACACCCGCGCAGTGGCGGCTGCTGGTGGCC 415
 QY 402 GGCCAACTACAGCAAGTGGAGCTGGGGCGGCACTGGCCCTGCTGGCCCTGGCCCTCACC 461
 DB 416 GGCCAACTACAGCAAGTGGAGCTGGGGCGGCACTGGCCCTGCTGGCCCTGGCCCTCACC 475
 QY 462 CGCCAGCTGGGGCCCGCGGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 521
 DB 476 CGCCAGCTGGGGCCCGCGGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 535
 QY 522 GCACGGACCGGCTGCTGGGGCCACCGGCTGGGGAGAC 558
 DB 536 GCACGGACCGGCTGCTGGGGCCACCGGCTGGGGAGAC 572
 RESULT 10
 LOCUS BD126485 670 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD126485
 VERSION BD126485.1 GI:23221430
 KEYWORDS JP 2002017375-A/1916.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 670)
 Ota.T., Nishikawa.T., Isogai.T., Hayashi.K., Ishii.S., Kawai.Y.,
 Wakamatsu.A., Sugiyama.T., Nagai.K., Kojima.S., Otsuki.T. and
 Koga.H.
 TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL HELIX RESEARCH INSTITUTE
 COMMENT OS Homo sapiens (human)
 PN JP 2002017375-A/1916
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC
 C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
 10, C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source 1..670
 FT /organism='Homo sapiens (human)'.
 FEATURES
 source Location/Qualifiers
 1..670
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'
 ORIGIN
 Query Match 31.4%; Score 346; DB 6; Length 670;
 Best Local Similarity 99.7%; Pred. No. 6.3e-171;
 Matches 396; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 162 GGGGGGCTCAAAACGCGAGCGCGGACCTGGCCCTGGCAAGTGGAGCTGACCATGGAGG 221
 DB 176 GGGGGGCTCAAAACGCGAGCGCGGACCTGGCCCTGGCAAGTGGAGCTGACCATGGAGG 235
 QY 222 TGGCCACATCTGGGGGGCTCCCTCATCGCCCGCTCTGGGTCTCTCCGCTGCTCACTG 281
 DB 236 TGGCCACATCTGGGGGGCTCCCTCATCGCCCGCTCTGGGTCTCTCCGCTGCTCACTG 295
 QY 282 TTTTCATGCAATGGGAGCTTGGAGCGCGCGAGTGGTTCGCTACTGCTGGCGGTGCA 341
 DB 296 TTTTCATGCAATGGGAGCTTGGAGCGCGCGAGTGGTTCGCTACTGCTGGCGGTGCA 355
 QY 342 CTCCAGGACGGGGCCCTGGAGCGCGCGACACCCGCGCAGTGGCGGCTGCTGGTGGCC 401
 DB 356 CTCCAGGACGGGGCCCTGGAGCGCGCGACACCCGCGCAGTGGCGGCTGCTGGTGGCC 415
 QY 402 GGCCAACTACAGCAAGTGGAGCTGGGGCGGCACTGGCCCTGCTGGCCCTGGCCCTCACC 461
 DB 416 GGCCAACTACAGCAAGTGGAGCTGGGGCGGCACTGGCCCTGCTGGCCCTGGCCCTCACC 475
 QY 462 CGCCAGCTGGGGCCCGCGGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 521
 DB 476 CGCCAGCTGGGGCCCGCGGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 535
 QY 522 GCACGGACCGGCTGCTGGGGCCACCGGCTGGGGAGAC 558
 DB 536 GCACGGACCGGCTGCTGGGGCCACCGGCTGGGGAGAC 572
 RESULT 11
 LOCUS AC009088/c 127769 bp DNA linear PRI 29-MAR-2003
 DEFINITION Homo sapiens chromosome 16 clone RP11-388M20, complete sequence.
 ACCESSION AC009088
 VERSION AC009088.9 GI:29366934

```
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 127769)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
              Alamos National Laboratory.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE     2 (bases 1 to 127769)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (03-APR-2002) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE     3 (bases 1 to 127769)
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
JOURNAL      Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE     4 (bases 1 to 127769)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
              Alamos National Laboratory.
TITLE        Direct Submission
JOURNAL      Submitted (18-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
REFERENCE     5 (bases 1 to 127769)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
              Alamos National Laboratory.
TITLE        Direct Submission
JOURNAL      Submitted (29-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT      On Mar 29, 2003 this sequence version replaced gi:29029216.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center and Los Alamos
              National Laboratory
              www.shgc.stanford.edu
              Quality: Phrap Quality >=40 99.7% of Sequence;
              Estimated Total Number of Errors is 0.2;
              NOTE: This insert is not the entire sequence of the clone (entire
              sequence is 233.4Kb). It is clipped at the overlap with AC135050.
              The number of bases overlapped is 2575.
FEATURES             Location/Qualifiers
     source           1..127769
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="REF1-388M20"
ORIGIN
Query Match      24.5%; Score 270; DB 9; Length 127769;
Best Local Similarity 100.0%; Pred. No. 5.8e-131;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 302 TGGAGCCCGCGCGAGTGTGCGTACTGTCTGGGCGTGCATCCAGGACGGGCCCCCTGG 361
      |||||
Db 31579 TGGAGCCCGCGCGCGAGTGTGCGTACTGTCTGGGCGTGCATCCAGGACGGGCCCCCTGG 31520
QY 362 ACGCGCGCGACACCCGCGCGAGTGGCGCGCATCTGTGTGCGCGGCCAATACAGCCAAGTGG 421
      |||||
Db 31519 ACGCGCGCGACACCCGCGCGAGTGGCGCGCATCTGTGTGCGCGGCCAATACAGCCAAGTGG 31460
QY 422 AGCTGGGCGCGAGCTGGCCCTGTGGCGCTGGCTCACCAGCCAGCTGGGCCCCCGCGG 481
      |||||
Db 31459 AGCTGGGCGCGAGCTGGCCCTGTGGCGCTGGCTCACCAGCCAGCTGGGCCCCCGCGG 31400
QY 482 TGTGGCTGTCTGCTGCGCGCGCGCTCACACCGCTTCTGTGACGGGACCGCCCTGCTGGG 541
      |||||
Db 31399 TGTGGCTGTCTGCTGCGCGCGCGCTCACACCGCTTCTGTGACGGGACCGCCCTGCTGGG 31340
QY 542 CCACCGGCTGGGGAGACGTCCAGGAGGAG 571
```

```
Db 31339 CCACCGGCTGGGAGAGCTCCAGGAGGAG 31310
      |||||
RESULT 12
AC135044 195476 bp DNA linear HTG 05-OCT-2002
LOCUS Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
DEFINITION SEQUENCE, 22 unordered pieces.
ACCESSION AC135044
VERSION AC135044.1 GI:23505535
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 195476)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 195476)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
              Center: Joint Genome Institute
              Center Code: JGI
              Web site: http://www.jgi.doe.gov
              -----
              Project Information
              Center Project Name: 809609
              Center clone name: CITB-El_2551B20
              -----
              Summary Statistics
              Consensus quality: 171229 bases at least Q40
              Consensus quality: 182638 bases at least Q30
              Consensus quality: 188095 bases at least Q20
              Estimated insert size: 170000; agarose-fp estimation
              Estimated insert size: 193376; sum-of-contigs estimation
              Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
              Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.
              NOTE: This is a 'working draft' sequence. It currently
              * consists of 22 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              * 1 1202: contig of 1202 bp in length
              * 1203 1302: gap of unknown length
              * 1303 2468: contig of 1166 bp in length
              * 2469 2568: gap of unknown length
              * 2569 4077: contig of 1509 bp in length
              * 4078 4177: gap of unknown length
              * 4178 5772: contig of 1595 bp in length
              * 5773 7722: gap of unknown length
              * 7723 7822: gap of unknown length
              * 7823 8958: contig of 1136 bp in length
              * 8959 11052: contig of 1993 bp in length
              * 11052 13641: contig of 2490 bp in length
              * 13642 13741: gap of unknown length
              * 13742 17738: contig of 3997 bp in length
              * 17739 17838: gap of unknown length
              * 17839 22030: contig of 4192 bp in length
              * 22031 27002: contig of 4872 bp in length
              * 27003 32819: contig of 5717 bp in length
              * 27103
```

```
* 32820 32919: gap of unknown length
* 32920 39003: contig of 6084 bp in length
* 39004 39103: gap of unknown length
* 48370 48370: contig of 9267 bp in length
* 48371 48470: gap of unknown length
* 48471 59721: contig of 11251 bp in length
* 59722 59821: gap of unknown length
* 59822 72304: contig of 12483 bp in length
* 72305 72404: gap of unknown length
* 72405 85349: contig of 12945 bp in length
* 85350 85449: gap of unknown length
* 85450 101872: contig of 16423 bp in length
* 101873 101972: gap of unknown length
* 101973 117076: contig of 15104 bp in length
* 117077 117176: gap of unknown length
* 117177 138973: contig of 21797 bp in length
* 138974 139074: gap of unknown length
* 139074 150337: contig of 11263 bp in length
* 150337 150437: gap of unknown length
* 150437 150476: contig of 45040 bp in length.
FEATURES             Location/Qualifiers
     source          1..195476
     organism="Homo sapiens"
     mol_type="genomic DNA"
     db_xref="taxon:9606"
     chromosome="16"
     clone_lib="CalTech human BAC library D"

ORIGIN
Query Match          24.5%; Score 270; DB 2; Length 195476;
Best Local Similarity 100.0%; Pred. No. 5.6e-131;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 302 TGGAGCCGCGCGCGAGTGTGCGTACTGCTGGGGTGCATCCAGACGCGGCCCTGG 361
Db 27417 TGGAGCCGCGCGCGAGTGTGCGTACTGCTGGGGTGCATCCAGACGCGGCCCTGG 27476

Qy 362 ACGGGCGCACACCGCGCATGCGCGCATGCTGGTGGCGGCACTACAGCAAGTGG 421
Db 27477 ACGGGCGCACACCGCGCATGCGCGCATGCTGGTGGCGGCACTACAGCAAGTGG 27536

Qy 422 AGCTGGGCGCGACCTGCGCCCTGCTGCGCCCTGCTACCGCGCAGCTGGGCCCCCGCG 481
Db 27537 AGCTGGGCGCGACCTGCGCCCTGCTGCGCCCTGCTACCGCGCAGCTGGGCCCCCGCG 27596

Qy 482 TGTGGCCTGTGCTGCTGCGCGCGCTCACACCGCTTGTGTCAGCGGACCGCTGCTGG 541
Db 27597 TGTGGCCTGTGCTGCTGCGCGCGCTCACACCGCTTGTGTCAGCGGACCGCTGCTGG 27656

Qy 542 CCACCGCTGGGGAGACGCTCCAGGAGGCG 571
Db 27657 CCACCGCTGGGGAGACGCTCCAGGAGGCG 27686

RESULT 13
AC093520/c
LOCUS              AC093520      168064 bp      DNA      linear      PRI 19-MAR-2003
DEFINITION        Homo sapiens chromosome 16 clone RP11-120K18, complete sequence.
ACCESSION          AC093520
VERSION            AC093520.4  GI:29124069
KEYWORDS            HTG.
SOURCE              Homo sapiens (human)
ORGANISM           Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1 (bases 1 to 168064)
AUTHORS            DOE Joint Genome Institute, Stanford Human Genome Center and Los
                  Alamos National Laboratory.
TITLE              Direct Submission
JOURNAL            Unpublished
REFERENCE          2 (bases 1 to 168064)
AUTHORS            DOE Joint Genome Institute.
```

```
Direct Submission
Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 168064)
DOE Joint Genome Institute.
Direct Submission
Submitted (13-JUN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 168064)
DOE Joint Genome Institute.
Direct Submission
Submitted (07-JUL-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
5 (bases 1 to 168064)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Mar 19, 2003 this sequence version replaced gi:21702833.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.sngc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1..168064
     organism="Homo sapiens"
     mol_type="genomic DNA"
     db_xref="taxon:9606"
     chromosome="16"
     clone="RP11-120K18"

ORIGIN
Query Match          7.9%; Score 87; DB 9; Length 168064;
Best Local Similarity 100.0%; Pred. No. 2.5e-34;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 952 CATCAACGGCATCTCGGATCTCGCCAGATCCCTTTGAGGCCCTTGCCCATATATCA 1011
Db 141592 CATCAACGGCATCTCGGATCTCGCCAGATCCCTTTGAGGCCCTTGCCCATATATCA 141533

Qy 1012 GTAGGAGTCTCACTGGGACCAAAAGC 1038
Db 141532 GTAGGAGTCTCACTGGGACCAAAAGC 141506

RESULT 14
AX098193
LOCUS              AX098193      1796 bp      DNA      linear      PAT 30-MAR-2001
DEFINITION        Sequence 105 from Patent WO0118542.
ACCESSION          AX098193
VERSION            AX098193.1  GI:13515276
KEYWORDS
SOURCE              Homo sapiens (human)
ORGANISM           Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1
AUTHORS            Lee, J., Thompson, P. and Lillie, J.
TITLE              Identification, assessment, prevention, and therapy of ovarian
                  cancer
JOURNAL            Patent: WO 0118542-A 105 15-MAR-2001;
                  Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1796
     organism="Homo sapiens"
     mol_type="unassigned DNA"
     db_xref="taxon:9606"

ORIGIN
```

```
Query Match      6.1%; Score 67; DB 6; Length 1796;
Best Local Similarity 100.0%; Pred. No. 1.3e-23;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGCCCTGTCTCTGGCCATGCCCGAGAGGGGTCTCTGGGCTGGGCGACGTGGGGCT 60
    |||||
Db 134 GGGCCCTGTCTCTGGCCATGCCCGAGAGGGGTCTCTGGGCTGGGCGACGTGGGGCT 193

Qy 61 GTGGCCA 67
    |||||
Db 194 GTGGCCA 200

RESULT 15
BC001462
LOCUS      BC001462      1800 bp      mRNA      linear      PRI 04-OCT-2003
DEFINITION Homo sapiens protease, serine, 8 (proctasin), mRNA (cDNA clone
MGC:2133 IMAGE:3138532), complete cds.
ACCESSION BC001462
VERSION   BC001462.2 GI:33876244
KEYWORDS  MGC.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 1800)
AUTHORS   Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J.J., Haieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Uesdin, T.B., Toohy, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullighy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Pahey, J., Helton, E., Kettman, M., Madan, A., Young, A.C., Shevchenko, Y.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalley, D.E.,
Schneringer, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1800)
Strausberg, R.
Direct Submission
Submitted (12-DEC-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
On Aug 19, 2003 this sequence version replaced gi:12655206.
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lario, P., Legaapi, R.,
Maduro, O.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAL Plate: 4 Row: J Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21536453.

FEATURES             Location/Qualifiers
     source            1..1800
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="MGC:2133 IMAGE:3138532"
                        /tissue_type="Placenta, Choriocarcinoma"
                        /clone_lib="NIH_MGC_21"
                        /lab_host="DH10B-R"
                        /note="Vector: pOT7"
     gene              1..1800
                        /gene="PRSS8"
                        /note="synonym: PROCTASIN"
                        /db_xref="LocusID:5652"
                        /db_xref="MIM:600823"
                        176..1207
                        /codon_start=1
                        /product="proctasin preproprotein"
                        /protein_id="AAH01462.1"
                        /db_xref="GI:12655207"
                        /db_xref="LocusID:5652"
                        /translation="MAQKGVLPQGLGAVAILLYLLRSGTGAEGAEPCGVAPQAR
ITGSSAVAGQWPQVTSITYGVHVCQSGSLVSEQWLSAAHCFSEHKEAYEVKGA
HOLDYSADAKVSTLKDIIIPHSYLOQSGDIALLOLRPTIPFSYRITCLPAANA
SFPNGHLCVTWGHVAPSLSLTPELQLEVLISRTCNCLYNIDAKPEEPHFVQ
EDMWACGYVEGGKDACGDSGGHLSCEVEGLWLTGLYVSGWDAGGANNRPQVYTLAS
YASWIOSKVTQLQPRVVVPTQESOPDNLCSHLAFSSAPAQGLRILFLPLGLALG
LLSPFLSEH"
     misc_feature      305..1018
                        /note="Tryp_SPC; Region: Trypsin-like serine protease"
                        /db_xref="CDD:smart00020"

ORIGIN
Query Match      6.1%; Score 67; DB 9; Length 1800;
Best Local Similarity 100.0%; Pred. No. 1.3e-23;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGCCCTGTCTCTGGCCATGCCCGAGAGGGGTCTCTGGGCTGGGCGACGTGGGGCT 60
    |||||
Db 158 GGGCCCTGTCTCTGGCCATGCCCGAGAGGGGTCTCTGGGCTGGGCGACGTGGGGCT 217

Qy 61 GTGGCCA 67
    |||||
Db 218 GTGGCCA 224

Search completed: February 25, 2004, 13:48:54
Job time : 4330 secs
```


PR 02-MAR-2001; 2001US-0272817P.
PR 05-JUL-2001; 2001US-030231P.
PR 12-JUL-2001; 2001US-0305060P.
PR 10-SEP-2001; 2001US-0318405P.
PR 12-SEP-2001; 2001US-0318700P.
PR 04-JAN-2002; 2002US-00037417.
XX (CURA-) CURAGEN CORP.
XX Kekuda R, Alsobrook JP, Tchernev VT, Liu X, Spytek KA;
PI Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CM, Li L;
PI Gorman L, Edinger S, Sciore P, Ellerman K, Malyankar U;
PI Rothenberg M, Stone D, Boldog F, Guo X, Shency S, Anderson D;
PI Padigaru W, Taupier RJ, Miller CE, Eisen A;
XX WPI; 2002-583619/62.
DR P-PSDB; ABB09523.
XX Novel polypeptides and nucleic acids homologous to transmembrane
XX receptor, thymosin, neuromodulin-like family of proteins for diagnosing,
PT treating cancer, atherosclerosis, neurological, skin and autoimmune
PT disorders.
XX
XX Claim 9a; Page 142; 323pp; English.
XX
XX The invention relates to 24 novel human proteins designated NOV1-NOV14
XX (ABB09501-ABB09524), collectively referred to as NOVX proteins, and
XX nucleic acids encoding them (ABQ33879-ABQ3902). NOVX proteins and
XX nucleotides are useful in the treatment, diagnosis or prevention of NOVX-
XX associated disorders or in the manufacture of a medicament for treating
XX such disorders, with specific applications described for each of the 24
XX NOVX proteins, based on their homology to known proteins. Various
XX disorders are associated with NOVX proteins including neurological
XX disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
XX pain, behavioural disorders, addiction, tuberculous sclerosis, cancers
XX (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
XX (e.g., allergic and autoimmune diseases), myasthenia gravis, asthma,
XX various forms of arthritis, diabetes, thyroiditis, cardiovascular disease
XX (e.g., hypertension), reproductive disorders, endometriosis,
XX incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,
XX cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
XX disorders, obesity, bacterial infections and particularly cardiomyopathy,
XX atherosclerosis, cell signal processing-related disorders and disorders
XX of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
XX be used to identify cellular receptors or downstream effectors which
XX binds to a NOVX protein, and are also useful as targets for the
XX identification of small molecules that modulate or inhibit processes such
XX as neurogenesis, cell differentiation, cell motility, cellular
XX proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
XX nucleic acid sequences can be used to identify a cell or tissue type and
XX are useful as a source of primers or probes for forensic biology and for
XX identifying and cloning NOVX homologues in other cell types. Cells
XX comprising NOVX nucleic acids are useful for producing non-human
XX transgenic animals which are useful for studying the function and
XX activity of NOVX proteins and for identifying and evaluating modulators
XX of NOVX activity. The present sequence represents DNA encoding the
XX prostatic precursor-like protein NOV14a. The gene encoding NOV14a is
XX located on chromosome 16
XX
XX Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 1102; DB 6; Length 1102;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GGGGCCCTTCTGGGCCATGGCCAGAGGGGGTCTGGGGCCCTGGGAGCTGGGGGCT 60
XX
XX 1 GGGGCCCTTCTGGGCCATGGCCAGAGGGGGTCTGGGGCCCTGGGAGCTGGGGGCT 60
XX
XX 61 GTGGCCAAATTCGACTCATCTACCTTACGGGTGGTGGCGCCCGACCCCTAGGGGC 120
XX
XX 61 GTGGCCAAATTCGACTCATCTACCTTACGGGTGGTGGCGCCCGACCCCTAGGGGC 120
XX

QY 121 CCCCCTACTGCGGCGCCCTGAGCCCTCGGCCCGCATGCTGGGGGGGCTCAAAACGCGAG 180
DB 121 CCCCCTACTGCGGCGCCCTGAGCCCTCGGCCCGCATGCTGGGGGGGCTCAAAACGCGAG 180
QY 181 CCGGGCACTGGCCCTTGGCAAGTGGAGCTGACCATGAGAGTGGCCACATCTGCGGGGGC 240
DB 181 CCGGGCACTGGCCCTTGGCAAGTGGAGCTGACCATGAGAGTGGCCACATCTGCGGGGGC 240
QY 241 TCCCTCATCGCCCTCTCTGGGTCTCTCCGTCTCACTGTTTTCATGACGAATGGGAG 300
DB 241 TCCCTCATCGCCCTCTCTGGGTCTCTCCGTCTCACTGTTTTCATGACGAATGGGAG 300
QY 301 TTGGAGCCCGCGCGAGTGGTGTCTGCTGGGGGTGCTCCAGAGCGGGCCCTG 360
DB 301 TTGGAGCCCGCGCGAGTGGTGTCTGCTGGGGGTGCTCCAGAGCGGGCCCTG 360
QY 361 GACGGCGGCACACCGCGGAGTGGCGGCATCTGCTGGTGGCGGCCAACTACAGCCAAAGTG 420
DB 361 GACGGCGGCACACCGCGGAGTGGCGGCATCTGCTGGTGGCGGCCAACTACAGCCAAAGTG 420
QY 421 GAGCTGGGCGCGAGACCTGGCCCTGCTGGCCCTGCTGCGCTGGCCCTGCGCCCGCCG 480
DB 421 GAGCTGGGCGCGAGACCTGGCCCTGCTGGCCCTGCTGCGCTGGCCCTGCGCCCGCCG 480
QY 481 GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
DB 481 GTGTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
QY 541 GGCACCGGCTGGGAGACCTGTCAGGAGGAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
DB 541 GGCACCGGCTGGGAGACCTGTCAGGAGGAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
QY 601 GAAGTGGAGCTAAGGCTGCTGGGCGGAGGCGACCTGTCATGCTCTTACAGCCAGCCGCT 660
DB 601 GAAGTGGAGCTAAGGCTGCTGGGCGGAGGCGACCTGTCATGCTCTTACAGCCAGCCGCT 660
QY 661 CCCTTCAACT 720
DB 661 CCCTTCAACT 720
QY 721 CGCAGGAGACCTGCGGAGGCTGCTGCTGGGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
DB 721 CGCAGGAGACCTGCGGAGGCTGCTGCTGGGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 781 TGGTTCCAGCAGGAATCACAGCTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 840
DB 781 TGGTTCCAGCAGGAATCACAGCTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 840
QY 841 GTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 841 GTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
QY 901 CTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB 901 CTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 961 GCATTCCTGATTTCTGCGAGATCTTTTGGGCGCTTTTGGGCGCTTTTGGGCGCTTTTGGGCG 1020
DB 961 GCATTCCTGATTTCTGCGAGATCTTTTGGGCGCTTTTGGGCGCTTTTGGGCGCTTTTGGGCG 1020
QY 1021 TCAACTGGGACCAAAAGCTTGTCT 1080
DB 1021 TCAACTGGGACCAAAAGCTTGTCT 1080
QY 1081 TGGGGGTTTCTGATGGGGCTTCC 1102
DB 1081 TGGGGGTTTCTGATGGGGCTTCC 1102
XX
XX RESULT 2
XX ABQ3902
XX ID ABQ3902 standard; DNA; 1102 BP.
XX

Qy 788 AGCAGGATACACAGCTTTGGGTTTGGTGTGACGAGAAACCGCCCTGGAGTTTCA 847
 Db 770 AGCAGGATACACAGCTTTGGGTTTGGTGTGACGAGAAACCGCCCTGGAGTTTCA 829
 Qy 848 CTGCTGTGGCTACCTATGAGCATGATACGGGAGCAGGTGATGGTTCAGAGCTGGGC 907
 Db 830 CTGCTGTGGCTACCTATGAGCATGATACGGGAGCAGGTGATGGTTCAGAGCTGGGC 889
 Qy 908 CTGCTTTCCACCCAGCCCGAGAGACCCAGTCAGAT 945
 Db 890 CTGCTTTCCACCCAGCCCGAGAGACCCAGTCAGAT 927

RESULT 4
 ABQ75956
 ID ABQ75956 standard; cDNA; 2681 BP.
 AC
 XX
 AC ABQ75956;
 XX
 DT 17-OCT-2002 (first entry)
 XX
 DE Human PMM encoding sequence Incyte ID 2751509CB1.
 XX
 KW Human; PMM; protein modification and maintenance molecule;
 KW anticonvulsant; neuroprotective; nootropic; cytosolic; antipsoriatic;
 KW antiasthmatic; dermatological; antidiabetic; antiparkinsonian;
 KW antianemic; antinflamatory; antitumor; antitubercular; cardiac;
 KW hepatotropic; osteopathic; antiemetic; antipruritic; virucide;
 KW antibacterial; fungicide; gastrointestinal; antidiarrheic; laxative;
 KW haemostatic; anti-HIV; thrombolytic; anticoagulant; gynaecological;
 KW cardiovascular; antiarteriosclerotic; hypotensive; vasotropic;
 KW antitumor; antirheumatic; immunosuppressive; antiallergic; antithyroid;
 KW nephrotropic; angiot; thyromimetic; antiarthritic; uropathic;
 KW ophthalmological; antiparasitic; tranquiliser; vulnerary; keratolytic;
 KW auditory; antiseborrheic; antidepressant; neuroleptic; antiinfertility;
 KW inflamatory; protozoacide; Crohn's disease; hypertension; autoimmune;
 KW scabies; neurological; Alzheimer's disease; reproductive;
 KW ectopic pregnancy; gene therapy; vaccine; disorder; prostatic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 EH Key Location/Qualifiers
 FT 1..2457
 FT /*tag= a
 FT /product= "protein modification and maintenance molecule"
 XX
 PN WO200246383-A2.
 XX
 PD 13-JUN-2002.
 XX
 PF 05-DEC-2001; 2001WO-US046964.
 XX
 PR 08-DEC-2000; 2000US-0254399P.
 PR 21-DEC-2000; 2000US-0257803P.
 PR 05-JAN-2001; 2001US-0260110P.
 PR 19-JAN-2001; 2001US-0262851P.
 PR 25-JAN-2001; 2001US-0264623P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Yue H, Azimzai Y, Kallick DA, Baughn MR, Griffin JA, Swarnakar A;
 PI Lal FG, Wallia NK, Hafalia AJA, Gandhi AR, Au-Young J, Elliott VS;
 PI Ramkumar J, Thangavelu K, Lu Y, Warren BA, Lu DM, Lee EA;
 PI Tribouley CM, Arvizu C, Deleage AM, Yao MG, Khan FA;
 PI Sanjanwala NM;
 XX
 DR WPI: 2002-519664/55.
 DR P-PSDB; ABB98135.
 XX
 PT New isolated Protein Modification and Maintenance polypeptides, useful
 for diagnosis, and treatment of e.g. gastrointestinal disorders.
 XX

PS Claim 5; Page 196; 200pp; English.
 XX
 CC The invention relates to an isolated Protein Modification and Maintenance
 (PMM) polypeptide. Polypeptides of the invention may be used in the
 CC diagnosis, treatment and prevention of disorders associated with
 CC decreased expression or activity of PMM. These include gastrointestinal
 CC disorders (e.g. Crohn's disease), cardiovascular disorders (e.g.
 CC hypertension), autoimmune/inflammatory disorders (e.g. anaemia), cell
 CC proliferative disorders, developmental disorders, epithelial disorders
 CC (e.g. scabies), neurological disorders (e.g. Alzheimer's disease)
 CC reproductive disorders (e.g. ectopic pregnancy), and in gene therapy or a
 CC vaccine for such diseases. They may also be used in the assessment of the
 CC effects of exogenous compound on the expression of nucleic acid and amino
 CC acid sequences of protein modification and maintenance molecules. The
 CC current sequence represents a human PMM encoding sequence of the
 CC invention, encoding a polypeptide which has been found to have homology
 CC with rat prolactin.
 XX
 SQ Sequence 2681 BP; 431 A; 919 C; 861 G; 470 T; 0 U; 0 Other;
 Query Match 65.0%; Score 716; DB 6; Length 2681;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 816; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 128 ACTCGGGCGCCCTGAGCCCTCGCCCGCATCTGTTGGGGGCTCAAAACGCGAGCGGGCA 187
 Db 110 ACTCGGGCGCCCTGAGCCCTCGCCCGCATCTGTTGGGGGCTCAAAACGCGAGCGGGCA 169
 Qy 188 CTTGGCCCTTGGCAGTGGAGCTGACCATGAGGTGGCCACATCTGGGGGGTCCCTCA 247
 Db 170 CTTGGCCCTTGGCAGTGGAGCTGACCATGAGGTGGCCACATCTGGGGGGTCCCTCA 229
 Qy 248 TCGCCCTCTCTGGGCTCTCTCCGCTGCTCACTGTTTCATGACGAATGGAGCTGGAGC 307
 Db 230 TCGCCCTCTCTGGGCTCTCTCCGCTGCTCACTGTTTCATGACGAATGGAGCTGGAGC 289
 Qy 308 CCGCGCGCGAGTGTGCTGATCTGTTGGGCTGCTCACTCCAGGACGGGCGCTGGAGCGG 367
 Db 290 CCGCGCGCGAGTGTGCTGATCTGTTGGGCTGCTCACTCCAGGACGGGCGCTGGAGCGG 349
 Qy 368 CGCACACCGCGCGAGTGGCCCATCTGTTGGGCTGCTCACTCCAGGACGGGCGCTGGAGCTGG 427
 Db 350 CGCACACCGCGCGAGTGGCCCATCTGTTGGGCTGCTCACTCCAGGACGGGCGCTGGAGCTGG 409
 Qy 428 GCGCGCGAGCTGGCCCTGCTGCGCTGCTCACTCCAGGACGGGCGCTGGAGCTGGG 487
 Db 410 GCGCGCGAGCTGGCCCTGCTGCGCTGCTCACTCCAGGACGGGCGCTGGAGCTGGG 469
 Qy 488 CTGCTGCTGCTGCGCGCTGCTCACTCCAGGACGGGCGCTGGAGCTGGGCGCTGGAGCTGG 547
 Db 470 CTGCTGCTGCTGCGCGCTGCTCACTCCAGGACGGGCGCTGGAGCTGGGCGCTGGAGCTGG 529
 Qy 548 GTTGGGGAGAGTCCAGGAGGAGATCTCTGCTGCTCTCCCTGGGCTGTCTACAGGAGTGG 607
 Db 530 GTTGGGGAGAGTCCAGGAGGAGATCTCTGCTGCTCTCCCTGGGCTGTCTACAGGAGTGG 589
 Qy 608 AGCTAAGGCTGTGGGCGAGGCGCACTGTCAATGTCTTACAGGACGGGCGCTGGAGCTGG 667
 Db 590 AGCTAAGGCTGTGGGCGAGGCGCACTGTCAATGTCTTACAGGACGGGCGCTGGAGCTGG 649
 Qy 668 ACTCACTCTCAGATATTCAGGAGGATGCTGTGTGGCTGTCTACAGGAGGCGCGAGG 727
 Db 650 ACTCACTCTCAGATATTCAGGAGGATGCTGTGTGGCTGTCTACAGGAGGCGCGAGG 709
 Qy 728 ACACCTGCGAGGCTGCTGCGGGCGCCCTGCTGTGTGAGGAGGCGGCGCTGGTTC 787
 Db 710 ACACCTGCGAGGCTGCTGCGGGCGCCCTGCTGTGTGAGGAGGCGGCGCTGGTTC 769
 Qy 788 AGGCGAGGATACACAGCTTTGGGTTTGGCTGTGTGAGGAGGAGAACCCCTGTGAGTTTCA 847
 Db 770 AGGCGAGGATACACAGCTTTGGGTTTGGCTGTGTGAGGAGGAGAACCCCTGTGAGTTTCA 829
 Qy 848 CTGCTGTGGCTACCTATGAGGATGATACGGGAGCAGGTGATGGTTCAGAGCTGGGC 907

CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 597 BP; 97 A; 217 C; 196 G; 87 T; 0 U; 0 Other;
Query Match 36.1%; Score 398; DB 5; Length 597;
Best Local Similarity 100.0%; Pred. No. 1.3e-175;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 341 ACTCCAGAGCGGCGCTGGAGCGGCGGACACCGCGCAGTGGCGGCATCGTGTGC 400
Db |||||
QY 401 CGGCCAATACACCAAGTGGAGCTGGCGCGCGACCTGGCCCTGCTGGCCTCTCAC 460
Db |||||
QY 110 CGGCCAATACACCAAGTGGAGCTGGCGCGCGACCTGGCCCTGCTGGCCTCTCAC 169
QY 461 CGGCCAATACACCAAGTGGAGCTGGCGCGCGACCTGGCCCTGCTGGCCTCTCAC 520
Db |||||
QY 170 CGGCCAATACACCAAGTGGAGCTGGCGCGCGACCTGGCCCTGCTGGCCTCTCAC 229
QY 521 TGCACGGCAGCGGCGCTGGAGCGGCGGACACCGCGCAGTGGCGGCATCGTGTGC 580
Db |||||
QY 230 TGCACGGCAGCGGCGCTGGAGCGGCGGACACCGCGCAGTGGCGGCATCGTGTGC 289
QY 581 CTCTCCCTGGGTGCTACAGAGTGGAGCTAAAGCTGTGGCGAGCGCCACCTGTCAAT 640
Db |||||
QY 290 CTCTCCCTGGGTGCTACAGAGTGGAGCTAAAGCTGTGGCGAGCGCCACCTGTCAAT 349
QY 641 GTCTCTACAGCGAGCGGCGCTCCCTTCAACCTCACTCTCCAGATATTGCCAGGATGTGT 700
Db |||||
QY 350 GTCTCTACAGCGAGCGGCGCTCCCTTCAACCTCACTCTCCAGATATTGCCAGGATGTGT 409
QY 701 GTCTCTACAGCGAGCGGCGCTCCCTTCAACCTCACTCTCCAGATATTGCCAGGATGTGT 738
Db |||||
QY 410 GTCTCTACAGCGAGCGGCGCTCCCTTCAACCTCACTCTCCAGATATTGCCAGGATGTGT 447

RESULT 8
ID ABK13566/c
XX ABK13566 standard; cDNA; 456 BP.
XX AC ABK13566;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human prostatic-like serine protease cDNA #2.
XX
KW Human; prostatic-like serine protease; cytostatic; antiatherosclerotic;
KW virucide; osteoparic; antiinflammatory; vasotropic; neuroprotective;
KW trypsin-like; metastasis; autoimmune lesion; atherosclerosis; gene;
KW renal failure; malignant cell; tumour angiogenesis; enzyme; inflammation;
KW chronic obstructive pulmonary disease; COPD; restenosis; atherosclerosis;
KW neurodegenerative disease; prion protein; infection; amyloid plaque;
KW Genstmann-Straussler Syndrome; viral infection; Scrapie;
KW Creutzfeldt-Jakob disease; metastatic cancer; lipid accumulation;
KW osteoporosis; Paget's disease; ss; Est; expressed sequence tag.
XX
OS Homo sapiens.
XX
PN W0200198467-A2.
XX
PD 27-DEC-2001.
XX
PF 22-JUN-2001; 2001WO-EP007117.
XX

PR 23-JUN-2000; 2000US-0213588P.
PR 20-MAR-2001; 2001US-0276909P.
XX (FARB) BAYER AG.
XX
XX Xiao Y, Morozov V;
XX WPI; 2002-114576/15.
XX
PT Novel human prostatic-like serine protease polypeptide and polynucleotide
PT which can be regulated for treating metastasis of malignant cells,
PT inflammation, atherosclerosis, neurodegenerative disease and infections.
XX
XX Disclosure; Fig 3; 11lpp; English.
XX
CC This invention comprises the cDNA and protein sequences of an isolated
CC prostatic-like serine protease and reagents and methods for regulating
CC the human prostatic-like enzyme activity. Prostatic is a trypsin-like
CC serine protease purified from human seminal fluid. An antibody specific
CC for prostatic-like serine protease is useful for immunodetection and
CC diagnosis of micro-metastases, autoimmune lesions and renal failure in
CC biopsy specimens, plasma samples and body fluids. The antibody may be
CC used to modulate enzyme activity in a disease, such as metastasis of
CC malignant cells, tumour angiogenesis, inflammation, chronic obstructive
CC pulmonary disease (COPD), atherosclerosis, neurodegenerative disease
CC (e.g. prion protein amyloid plaques of Genstmann-Straussler Syndrome,
CC Creutzfeldt-Jakob disease, Scrapie) or infection, particularly viral
CC infection. The human prostatic-like serine protease gene provides a
CC therapeutic target of decreasing human prostatic-like serine protease
CC activity, in particular for treating or preventing metastatic cancer. The
CC agonists and antagonists of the nucleotide sequence may be used to mimic,
CC augment and inhibit the enzyme activity which may be useful to treat
CC osteoporosis, Paget's disease and degradation of bone implants
CC particularly dental implants. Altered levels of human prostatic-like
CC serine protease activity inhibits both smooth muscle cell proliferation
CC and lipid accumulation and inhibit the progression of restenosis and
CC atherosclerosis. The nucleic acid sequence is also useful in diagnostic
CC assays for detecting diseases and abnormalities or susceptibility to
CC diseases related to the presence of mutations in nucleic acid sequences
CC which encode the enzyme. The present sequence represents the human
CC prostatic-like serine protease #2 nucleotide sequence of the invention
XX
SQ Sequence 456 BP; 74 A; 151 C; 167 G; 62 T; 0 U; 2 Other;
Query Match 34.8%; Score 384; DB 6; Length 456;
Best Local Similarity 100.0%; Pred. No. 4.7e-169;
Matches 384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 379 GCAGTGGCGCGCCATCGTGTGTCGCGGCCCACTACAGCCAACTGGAGCTGGCGCGACCTG 438
Db |||||
QY 384 GCAGTGGCGCGCCATCGTGTGTCGCGGCCCACTACAGCCAACTGGAGCTGGCGCGACCTG 325
QY 439 GCGCTGTGTGCGCTGTGCTTACCGCGCAGCTGGCGCGCGCTGTGCTGTGCTGTGCTGTG 498
Db |||||
QY 324 GCGCTGTGTGCGCTGTGCTTACCGCGCAGCTGGCGCGCGCTGTGCTGTGCTGTGCTGTG 265
QY 499 CCGCGCGCTTACAGCCGCTTGTGTCAGCGCACCGCCCTGTGGGCCACCGGCTGGGGAGAC 558
Db |||||
QY 264 CCGCGCGCTTACAGCCGCTTGTGTGTCAGCGCACCGCCCTGTGGGCCACCGGCTGGGGAGAC 205
QY 559 GTCCAGGAGGCGAGATCTTCTGCTTCTCCCTGGGTGTGTACAGGAAGTGGAGCTTAAGCTG 618
Db |||||
QY 204 GTCCAGGAGGCGAGATCTTCTGCTTCTCCCTGGGTGTGTACAGGAAGTGGAGCTTAAGCTG 145
QY 619 CTGGCGGAGGCGGACCTGTCTATGTCTTACAGCCAGCCCGCTTCTTCACTCTCTCTCTC 678
Db |||||
QY 144 CTGGCGGAGGCGGACCTGTCTATGTCTTACAGCCAGCCCGCTTCTTCACTCTCTCTCTC 85
QY 679 CAGATATTGGCAGGAGTGTCTGTGTGCTTACCCAGAGGGCCGCGAGGACACTCTGCCAG 738
Db |||||
QY 84 CAGATATTGGCAGGAGTGTCTGTGTGCTTACCCAGAGGGCCGCGAGGACACTCTGCCAG 25
QY 739 GGTGACTCTGG 762

356	CTCCAGGACGGGCCCCTGGACGGCGCGACACCCGCGCAGTGGCGCGCCATCGTGGTGCC	415
402	GGCCAACTACAGCAACATGGAGCTGGCGCGCAGACCTGGGCCCTGCTGGCCTGGCCTCACC	461
416	GGCCAACTACAGCCCAATGGAGCTGGCGCGCAGACCTGGCCCTGCTGGCCTGGCCTCACC	475
462	CGCCAGCCTGGGGCCCCGCGCGTGTGGCTGTCTGCTGCCCGGGCCTCACACGGCTTCGT	521
476	CGCCAGCCTGGGGCCCCGCGCGTGTGGCTGTCTGCTGCCCGGGCCTCACACGGCTTCGT	535
522	GACGGCACCGCCTGTGGGGCCACCGGCTGGGGAGAC	558
536	GACGGCACCGCCTGTGGGGCCACCGGCTGGGGAGAC	572

RESULT 10

AAK93456
ID AAK93456 standard; cDNA; 670 BP.
XX
XX AAK93456;
XX AC
XX DT 06-NOV-2001 (first entry)
XX DE Human cDNA clone representative sequence, SEQ ID NO: 1916.
XX DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX KW
XX OS Homo sapiens.
XX OS
XX PN EP1130094-A2.
XX XX
XX PD 05-SEP-2001.
XX XX
XX PF 07-JUL-2000; 2000EP-00114089.
XX PF
XX PR 08-JUL-1999; 99JP-00194486.
XX PR 11-JAN-2000; 2000JP-0018774.
XX PR 02-MAY-2000; 2000JP-00183765.
XX XX
XX XX (HELIJ-) HELIX RES INST.

The invention relates to primers for synthesising full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5' and 3'-ends of the cDNA molecules have been determined. Primers for synthesising the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is the nucleotide sequence of the 5'-end of a cDNA provided in the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO

sequence 570 bp: 97 A: 245 C: 205 G: 118 T: 0 U: 5 Other:

	Query Match	31.4%;	Score 346;	DB 4;	Length 670;	
	Best Local Similarity	99.7%;	Pred. NO. 2.6e-151;			
	Matches	396;	Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;
QY	162	GGGGGGCTAAACGCGCAGCGGGGCACCTGGCCCTTGGCAAGTGAAGCTGCACCATGGAGG	221			
Db	176	GGGGGGCTAAACGCGCAGCGGGGCACCTGGCCCTTGGCAAGTGAAGCTGCACCATGGAGG	235			
QY	222	TGGGCA CATCTGGGGGGCTCCCTATCGCCCCCTCTCTGGGTCTCTTCGGCTGCTCACTG	281			
Db	236	TGGGCA CATCTGGGGGGCTCCCTATCGCCCCCTCTCTGGGTCTCTTCGGCTGCTCACTG	295			
QY	282	TTTTCATGACGAATGGGACGTTTGGAGCCCGCGCCAGTGTGGTACCTGTGGGGGTGCA	341			
Db	296	TTTTCATGACGAATGGGACGTTTGGAGCCCGCGCCAGTGTGGTACCTGTGGGGGTGCA	355			
QY	342	CTCCAGGACGGGGCCCTTGGACGGCGCGCACACCCGCGCACTGTGGTGGCC	401			

Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI; 2001-524255/58.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their use
XX in genetic manipulation.
XX
XX Example 11; SEQ ID NO 1916; 1380pp + Sequence Listing; English.
XX
XX The invention relates to primers for synthesising full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been isolated
XX and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
XX been determined. Primers for synthesising the full length cDNA are useful
XX for clarifying the function of the protein encoded by the cDNA. The full
XX length clones were obtained by construction of full length enriched cDNA
XX libraries that were synthesised by the oligo-capping method. The primers
XX enable the production of the full length cDNA easily without any special
XX methods. The present sequence was used as the representative sequence
XX from a human clone which was used in homology searches to identify the
XX clone. Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in CD-ROM format directly from
XX EPO

```

XX
SQ      Sequence 670 BP; 97 A; 245 C; 205 G; 118 T; 0 U; 5 Other;
        31.4%; Score 346; DB 4; Length 670;
Query Match
Best Local Similarity 99.7%; Pred. No. 2.6e-151;
Matches 396; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      162 GGGGGGCTCAACGGCGAGCGCGGCACTGGCCCTTGGCAACTGAGCTGCACCATGAGG 221
        |||||

```


PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

PS Claim 1; SEQ ID NO 6883; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2298 BP; 468 A; 774 C; 706 G; 350 T; 0 U; 0 Other;

Query Match 12.3%; Score 135; DB 5; Length 2298;
 Best Local Similarity 100.0%; Pred. No. 8.6e-53;
 Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 811 TTTGGCTGTGACGAGAAACCGCCCTGGAGTTTTCATCTGTGGCTACCTATGAGGCA 870
 DB 445 TTTGGCTGTGACGAGAAACCGCCCTGGAGTTTTCATCTGTGGCTACCTATGAGGCA 504
 QY 871 TGGATACGGGAGAGGTGATGGTTTCAGAGCTGGGCTGCTTCCACCCAGCCCCAG 930
 DB 505 TGGATACGGGAGAGGTGATGGTTTCAGAGCTGGGCTGCTTCCACCCAGCCCCAG 564
 QY 931 AAGACCCAGTCAGAT 945
 DB 565 AAGACCCAGTCAGAT 579

RESULT 13
 APT31936
 ID APT31936 standard; DNA; 1733 BP.

XX APT31936;
 XX 01-MAY-2003 (first entry)
 XX Human breast cancer / ovarian cancer related coding sequence #43.
 DE Human; gene; ds; cytostatic; breast cancer; ovarian cancer.
 KW Homo sapiens.
 OS WO2003000012-A2.

XX 03-JAN-2003.
 XX 21-JUN-2002; 2002WO-US019773.
 XX 21-JUN-2001; 2001US-0300159P.
 XX 27-JUN-2001; 2001US-0301351P.
 XX (MILL-) MILLENNIUM PHARM INC.

PI Veiby OP;
 XX WPI; 2003-267848/26.
 DR P-PSDB; ABJ37067.

XX Determining the presence of breast cancer in an individual, involves
 PT using specific polynucleotide markers.

PS Disclosure; Page 192-193; 233pp; English.

XX The invention comprises a method for assessing whether a patient is
 CC afflicted with breast cancer or ovarian cancer. The method involves the
 CC use of specific DNA markers. The method of the invention is useful in the
 CC detection and treatment of ovarian and breast cancer. DNA sequences
 CC APT31894 - APT31949 encode human breast/ovarian cancer-related proteins
 XX Sequence 1733 BP; 311 A; 578 C; 500 G; 344 T; 0 U; 0 Other;

Query Match 6.1%; Score 67; DB 7; Length 1733;
 Best Local Similarity 100.0%; Pred. No. 5e-21;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCCCTTGTCTGGGCGCATGGCCCAAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 60
 DB 81 GGACCCCTTGTCTGGGCGCATGGCCCAAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 140
 QY 61 GTGGCCA 67
 DB 141 GTGGCCA 147

RESULT 14
 AAF98698
 ID AAF98698 standard; DNA; 1796 BP.

XX AAF98698;
 XX 02-JUL-2001 (first entry)
 XX Human ovarian cancer cell expressed sequence 10798.
 DE Human; ovarian cancer; identification; detection; characterisation;
 KW tumour; kinase; marker; cytostatic; antisense gene therapy; ds.

OS Homo sapiens.
 XX WO200118542-A2.
 XX 15-MAR-2001.

XX 01-SEP-2000; 2000WO-US024199.
 XX 03-SEP-1999; 99US-0152547P.
 XX 16-MAR-2000; 2000US-0190347P.
 XX 21-MAR-2000; 2000US-0191321P.
 XX 31-MAY-2000; 2000US-0208382P.
 XX 20-JUL-2000; 2000US-00220467.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lee J, Thomsho P, Lillie J;
 WPI; 2001-211428/21.

PT Detection, assessment, prevention and therapy of ovarian cancer,
 PT comprises detecting changes in the expression of a variety of markers.

PS Claim 1; Page 1001-1002; 1198pp; English.

XX The present invention describes a method for assessing whether a patient
 CC is afflicted with ovarian cancer by comparing: (1) the expression of a
 CC marker (I) (see AAF98594 to AAF98730), in a patient sample; and (2) the
 CC normal level of expression of (I) in a control non-ovarian cancer sample,

CC where a significant difference between the level of expression in (a) and
CC (b) is an indication that the patient is afflicted with ovarian cancer.
CC (I) have cytostatic activities and can be used in antisense gene therapy.
CC The method, compositions and kits from the present invention can be used
CC for: (1) assessing and treating ovarian cancer; (2) making isolated
CC hybridoma, which produces an antibody useful for ovarian cancer
CC assessment; and (3) inhibiting ovarian cancer in a patient. AAF98573 to
CC AAF98593 represent human kinase marker primers and probes which are used
CC in the exemplification of the present invention

XX
SQ Sequence 1796 BP; 327 A; 599 C; 511 G; 359 T; 0 U; 0 Other;

Query Match 6.1%; Score 67; DB 5; Length 1796;
Best Local Similarity 100.0%; Pred. No. 5e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 60
|||
Db 134 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 193
|||
QY 61 GTGGCCA 67
|||
Db 194 GTGGCCA 200
|||

RESULT 15
AAF98720
ID AAF98720 standard; DNA; 1835 BP.

XX
AC AAF98720;

XX
DT 02-JUL-2001 (first entry)

XX
DE Human late stage ovarian tumour polynucleotide marker 28.

XX
KW Human; ovarian cancer; identification; detection; characterisation;
KW tumour; kinase; marker; cytostatic; antisense gene therapy; ds.

XX
OS Homo sapiens.

XX
FN WO200118542-A2.

XX
PD 15-MAR-2001.

XX
PF 01-SEP-2000; 2000WO-US024199.

XX
PR 03-SEP-1999; 99US-0152547P.

XX
PR 16-MAR-2000; 2000US-0190347P.

XX
PR 21-MAR-2000; 2000US-0191321P.

XX
PR 31-MAY-2000; 2000US-0208382P.

XX
PR 20-JUL-2000; 2000US-00220467.

XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX
PI Lee J, Thompsho P, Lillie J;

XX
PT WPI; 2001-211428/21.

XX
PS Claim 1; Page 1186-1187; 1198pp; English.

XX
CC The present invention describes a method for assessing whether a patient
CC is afflicted with ovarian cancer by comparing: (1) the expression of a
CC marker (I) (see AAF98594 to AAF98730), in a patient sample; and (2) the
CC normal level of expression of (I) in a control non-ovarian cancer sample,
CC where a significant difference between the level of expression in (a) and
CC (b) is an indication that the patient is afflicted with ovarian cancer.
CC (I) have cytostatic activities and can be used in antisense gene therapy.
CC The method, compositions and kits from the present invention can be used
CC for: (1) assessing and treating ovarian cancer; (2) making isolated
CC hybridoma, which produces an antibody useful for ovarian cancer

CC assessment; and (3) inhibiting ovarian cancer in a patient. AAF98573 to
CC AAF98593 represent human kinase marker primers and probes which are used
CC in the exemplification of the present invention

XX
SQ Sequence 1835 BP; 309 A; 621 C; 527 G; 378 T; 0 U; 0 Other;

Query Match 6.1%; Score 67; DB 5; Length 1835;
Best Local Similarity 100.0%; Pred. No. 5e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 60
|||
Db 212 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 271
|||
QY 61 GTGGCCA 67
|||
Db 272 GTGGCCA 278
|||

Search completed: February 25, 2004, 12:36:33
Job time : 478 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 12:18:11 ; Search time 2783 Seconds
(without alignments)
11824.694 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102
Sequence: 1 gggcccttctctgggcccatt.....ggggttctgatggggcctcc 1102

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pig:*
27: em_gss_vri:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	466	42.3	537	10	AW450407
C 2	384	34.8	456	9	AI190509
C 3	135	12.3	930	13	BX436299
4	67	6.1	294	9	AA300017

67	6.1	318	12	BM828821
67	6.1	360	9	AI393077
67	6.1	366	9	AI623099
67	6.1	377	9	AI193435
67	6.1	406	14	CB116948
67	6.1	409	9	AU298208
67	6.1	462	9	AI343968
67	6.1	472	9	AI761519
67	6.1	479	9	AU298317
67	6.1	479	14	CB996389
67	6.1	492	10	BF063417
67	6.1	516	12	BM311864
67	6.1	527	9	AI658798
67	6.1	527	12	BG482000
67	6.1	534	9	AI393068
67	6.1	537	12	BM837284
67	6.1	537	12	BM837397
67	6.1	542	10	BE280845
67	6.1	555	12	BM837589
67	6.1	570	12	BG479276
67	6.1	570	14	CB125988
67	6.1	575	9	AI207222
67	6.1	583	12	BG386531
67	6.1	590	13	BQ292219
67	6.1	591	9	AU134797
67	6.1	594	10	BE281529
67	6.1	604	9	AU298212
67	6.1	608	14	CD723955
67	6.1	618	9	AU135364
67	6.1	619	12	BM794592
67	6.1	625	9	AU134318
67	6.1	632	9	AU133801
67	6.1	648	9	AU138370
67	6.1	657	9	AI935459
67	6.1	659	9	AU139645
67	6.1	666	9	AU136193
67	6.1	670	9	AU121662
67	6.1	681	9	AU139870
67	6.1	683	13	BX505123
67	6.1	688	9	AU136859
67	6.1	690	12	BG481221

ALIGNMENTS

RESULT 1
AW450407/c
LOCUS
DEFINITION
IMAGE:2735037 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AW450407 537 bp mRNA linear EST 17-FEB-2000
UI-H-BI3-akn-g-11-0-UI-s1 NCI CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2735037 3', mRNA sequence.
AW450407
AW450407.1 GI:6991183
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 537)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
is likely internal to the message. cDNA library preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
Information can be found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=No. Location/Qualifiers

BM828821 K-EST0101
AI393077 tg25f08.x
AI623099 tu49f05.x
AI193435 qe58e10.x
CB116948 K-EST0161
AU298208 AU298208
AI343968 tg01405.x
AI761519 wf16102.x
AU298317 AU298317
CB996389 AGENCOURT
BF063417 7h89e10.x
BM311864 1963a10.y
AI658798 tu08g09.x
BG482000 602527844
AI393068 tg25e09.x
BM837284 K-EST0113
BM837397 K-EST0113
BE280845 601155410
BM837589 K-EST0113
BG479276 602526495
CB125988 K-EST0174
AI207222 qf81h10.x
BG386531 602455771
BQ292219 PM2-AN008
AU134797 AU134797
BE281529 601155121
AU298212 AU298212
CD723955 OJ28e07.Y
AU135364 AU135364
BM794592 K-EST0075
AU134318 AU134318
AU138370 AU138370
AU133801 AU133801
AU935459 w084h02.x
AU139645 AU139645
AU136193 AU136193
AU121662 AU121662
AU139870 AU139870
BX505123 DKEZp686J
AU136859 AU136859
BG481221 602528547

```

source
1. .537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2735037"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_Sub5"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; NCI CGAP Sub5
is a subtracted library derived from NCI CGAP Sub4. The
NCI CGAP Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI CGAP Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI CGAP Kid5 pool 1 LLM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clonoids 1323912-1325831, 1471368-1472903,
1492104-1492255); NCI CGAP_Lu5 pool 1 LLM 3575-3582,
3851-3854 (IMAGE Clonoids
1414920-1417991, 1520904-1522439); NCI CGAP_G4 pool 1 LLM
3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids
1257096-1258631, 1469064-1470983, 1475592-1476743);
NCI CGAP_Pr22 pool 1 LLM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clonoids 985608-986759, 1101192-1101959,
1217928-1220615); NCI CGAP_Col0 pool 1 LLM 2644-2653,
2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (10% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI CGAP Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (10% of
the driver population), plus a pool of 11,136 clones from
NCI CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (10% of the
driver population), plus a pool of 5,472 clones from
NCI CGAP_Sub4 (IMAGE Clonoids 2723592-2728969) (70% of the
driver population). Subtraction was performed as
previously described (Bonald, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG TISSUE=colon
TAG_LIB=NCI CGAP_Col0
TAG_SEQ=AAACGG"

106 CTCTCCAGATATTGCCAGGATGCTGTGTGCTTACCCAGAGGGCGCAGGACACT 47
107 |||||
108 |||||
109 |||||
110 |||||
111 |||||
112 |||||
113 |||||
114 |||||
115 |||||
116 |||||
117 |||||
118 |||||
119 |||||
120 |||||
121 |||||
122 |||||
123 |||||
124 |||||
125 |||||
126 |||||
127 |||||
128 |||||
129 |||||
130 |||||
131 |||||
132 |||||
133 |||||
134 |||||
135 |||||
136 |||||
137 |||||
138 |||||
139 |||||
140 |||||
141 |||||
142 |||||
143 |||||
144 |||||
145 |||||
146 |||||
147 |||||
148 |||||
149 |||||
150 |||||
151 |||||
152 |||||
153 |||||
154 |||||
155 |||||
156 |||||
157 |||||
158 |||||
159 |||||
160 |||||
161 |||||
162 |||||
163 |||||
164 |||||
165 |||||
166 |||||
167 |||||
168 |||||
169 |||||
170 |||||
171 |||||
172 |||||
173 |||||
174 |||||
175 |||||
176 |||||
177 |||||
178 |||||
179 |||||
180 |||||
181 |||||
182 |||||
183 |||||
184 |||||
185 |||||
186 |||||
187 |||||
188 |||||
189 |||||
190 |||||
191 |||||
192 |||||
193 |||||
194 |||||
195 |||||
196 |||||
197 |||||
198 |||||
199 |||||
200 |||||
201 |||||
202 |||||
203 |||||
204 |||||
205 |||||
206 |||||
207 |||||
208 |||||
209 |||||
210 |||||
211 |||||
212 |||||
213 |||||
214 |||||
215 |||||
216 |||||
217 |||||
218 |||||
219 |||||
220 |||||
221 |||||
222 |||||
223 |||||
224 |||||
225 |||||
226 |||||
227 |||||
228 |||||
229 |||||
230 |||||
231 |||||
232 |||||
233 |||||
234 |||||
235 |||||
236 |||||
237 |||||
238 |||||
239 |||||
240 |||||
241 |||||
242 |||||
243 |||||
244 |||||
245 |||||
246 |||||
247 |||||
248 |||||
249 |||||
250 |||||
251 |||||
252 |||||
253 |||||
254 |||||
255 |||||
256 |||||
257 |||||
258 |||||
259 |||||
260 |||||
261 |||||
262 |||||
263 |||||
264 |||||
265 |||||
266 |||||
267 |||||
268 |||||
269 |||||
270 |||||
271 |||||
272 |||||
273 |||||
274 |||||
275 |||||
276 |||||
277 |||||
278 |||||
279 |||||
280 |||||
281 |||||
282 |||||
283 |||||
284 |||||
285 |||||
286 |||||
287 |||||
288 |||||
289 |||||
290 |||||
291 |||||
292 |||||
293 |||||
294 |||||
295 |||||
296 |||||
297 |||||
298 |||||
299 |||||
300 |||||
301 |||||
302 |||||
303 |||||
304 |||||
305 |||||
306 |||||
307 |||||
308 |||||
309 |||||
310 |||||
311 |||||
312 |||||
313 |||||
314 |||||
315 |||||
316 |||||
317 |||||
318 |||||
319 |||||
320 |||||
321 |||||
322 |||||
323 |||||
324 |||||
325 |||||
326 |||||
327 |||||
328 |||||
329 |||||
330 |||||
331 |||||
332 |||||
333 |||||
334 |||||
335 |||||
336 |||||
337 |||||
338 |||||
339 |||||
340 |||||
341 |||||
342 |||||
343 |||||
344 |||||
345 |||||
346 |||||
347 |||||
348 |||||
349 |||||
350 |||||
351 |||||
352 |||||
353 |||||
354 |||||
355 |||||
356 |||||
357 |||||
358 |||||
359 |||||
360 |||||
361 |||||
362 |||||
363 |||||
364 |||||
365 |||||
366 |||||
367 |||||
368 |||||
369 |||||
370 |||||
371 |||||
372 |||||
373 |||||
374 |||||
375 |||||
376 |||||
377 |||||
378 |||||
379 |||||
380 |||||
381 |||||
382 |||||
383 |||||
384 |||||
385 |||||
386 |||||
387 |||||
388 |||||
389 |||||
390 |||||
391 |||||
392 |||||
393 |||||
394 |||||
395 |||||
396 |||||
397 |||||
398 |||||
399 |||||
400 |||||
401 |||||
402 |||||
403 |||||
404 |||||
405 |||||
406 |||||
407 |||||
408 |||||
409 |||||
410 |||||
411 |||||
412 |||||
413 |||||
414 |||||
415 |||||
416 |||||
417 |||||
418 |||||
419 |||||
420 |||||
421 |||||
422 |||||
423 |||||
424 |||||
425 |||||
426 |||||
427 |||||
428 |||||
429 |||||
430 |||||
431 |||||
432 |||||
433 |||||
434 |||||
435 |||||
436 |||||
437 |||||
438 |||||
439 |||||
440 |||||
441 |||||
442 |||||
443 |||||
444 |||||
445 |||||
446 |||||
447 |||||
448 |||||
449 |||||
450 |||||
451 |||||
452 |||||
453 |||||
454 |||||
455 |||||
456 |||||
457 |||||
458 |||||
459 |||||
460 |||||
461 |||||
462 |||||
463 |||||
464 |||||
465 |||||
466 |||||
467 |||||
468 |||||
469 |||||
470 |||||
471 |||||
472 |||||
473 |||||
474 |||||
475 |||||
476 |||||
477 |||||
478 |||||
479 |||||
480 |||||
481 |||||
482 |||||
483 |||||
484 |||||
485 |||||
486 |||||
487 |||||
488 |||||
489 |||||
490 |||||
491 |||||
492 |||||
493 |||||
494 |||||
495 |||||
496 |||||
497 |||||
498 |||||
499 |||||
500 |||||
501 |||||
502 |||||
503 |||||
504 |||||
505 |||||
506 |||||
507 |||||
508 |||||
509 |||||
510 |||||
511 |||||
512 |||||
513 |||||
514 |||||
515 |||||
516 |||||
517 |||||
518 |||||
519 |||||
520 |||||
521 |||||
522 |||||
523 |||||
524 |||||
525 |||||
526 |||||
527 |||||
528 |||||
529 |||||
530 |||||
531 |||||
532 |||||
533 |||||
534 |||||
535 |||||
536 |||||
537 |||||
538 |||||
539 |||||
540 |||||
541 |||||
542 |||||
543 |||||
544 |||||
545 |||||
546 |||||
547 |||||
548 |||||
549 |||||
550 |||||
551 |||||
552 |||||
553 |||||
554 |||||
555 |||||
556 |||||
557 |||||
558 |||||
559 |||||
560 |||||
561 |||||
562 |||||
563 |||||
564 |||||
565 |||||
566 |||||
567 |||||
568 |||||
569 |||||
570 |||||
571 |||||
572 |||||
573 |||||
574 |||||
575 |||||
576 |||||
577 |||||
578 |||||
579 |||||
580 |||||
581 |||||
582 |||||
583 |||||
584 |||||
585 |||||
586 |||||
587 |||||
588 |||||
589 |||||
590 |||||
591 |||||
592 |||||
593 |||||
594 |||||
595 |||||
596 |||||
597 |||||
598 |||||
599 |||||
600 |||||
601 |||||
602 |||||
603 |||||
604 |||||
605 |||||
606 |||||
607 |||||
608 |||||
609 |||||
610 |||||
611 |||||
612 |||||
613 |||||
614 |||||
615 |||||
616 |||||
617 |||||
618 |||||
619 |||||
620 |||||
621 |||||
622 |||||
623 |||||
624 |||||
625 |||||
626 |||||
627 |||||
628 |||||
629 |||||
630 |||||
631 |||||
632 |||||
633 |||||
634 |||||
635 |||||
636 |||||
637 |||||
638 |||||
639 |||||
640 |||||
641 |||||
642 |||||
643 |||||
644 |||||
645 |||||
646 |||||
647 |||||
648 |||||
649 |||||
650 |||||
651 |||||
652 |||||
653 |||||
654 |||||
655 |||||
656 |||||
657 |||||
658 |||||
659 |||||
660 |||||
661 |||||
662 |||||
663 |||||
664 |||||
665 |||||
666 |||||
667 |||||
668 |||||
669 |||||
670 |||||
671 |||||
672 |||||
673 |||||
674 |||||
675 |||||
676 |||||
677 |||||
678 |||||
679 |||||
680 |||||
681 |||||
682 |||||
683 |||||
684 |||||
685 |||||
686 |||||
687 |||||
688 |||||
689 |||||
690 |||||
691 |||||
692 |||||
693 |||||
694 |||||
695 |||||
696 |||||
697 |||||
698 |||||
699 |||||
700 |||||
701 |||||
702 |||||
703 |||||
704 |||||
705 |||||
706 |||||
707 |||||
708 |||||
709 |||||
710 |||||
711 |||||
712 |||||
713 |||||
714 |||||
715 |||||
716 |||||
717 |||||
718 |||||
719 |||||
720 |||||
721 |||||
722 |||||
723 |||||
724 |||||
725 |||||
726 |||||
727 |||||
728 |||||
729 |||||
730 |||||
731 |||||
732 |||||
733 |||||
734 |||||
735 |||||
736 |||||
737 |||||
738 |||||
739 |||||
740 |||||
741 |||||
742 |||||
743 |||||
744 |||||
745 |||||
746 |||||
747 |||||
748 |||||
749 |||||
750 |||||
751 |||||
752 |||||
753 |||||
754 |||||
755 |||||
756 |||||
757 |||||
758 |||||
759 |||||
760 |||||
761 |||||
762 |||||
763 |||||
764 |||||
765 |||||
766 |||||
767 |||||
768 |||||
769 |||||
770 |||||
771 |||||
772 |||||
773 |||||
774 |||||
775 |||||
776 |||||
777 |||||
778 |||||
779 |||||
780 |||||
781 |||||
782 |||||
783 |||||
784 |||||
785 |||||
786 |||||
787 |||||
788 |||||
789 |||||
790 |||||
791 |||||
792 |||||
793 |||||
794 |||||
795 |||||
796 |||||
797 |||||
798 |||||
799 |||||
800 |||||
801 |||||
802 |||||
803 |||||
804 |||||
805 |||||
806 |||||
807 |||||
808 |||||
809 |||||
810 |||||
811 |||||
812 |||||
813 |||||
814 |||||
815 |||||
816 |||||
817 |||||
818 |||||
819 |||||
820 |||||
821 |||||
822 |||||
823 |||||
824 |||||
825 |||||
826 |||||
827 |||||
828 |||||
829 |||||
830 |||||
831 |||||
832 |||||
833 |||||
834 |||||
835 |||||
836 |||||
837 |||||
838 |||||
839 |||||
840 |||||
841 |||||
842 |||||
843 |||||
844 |||||
845 |||||
846 |||||
847 |||||
848 |||||
849 |||||
850 |||||
851 |||||
852 |||||
853 |||||
854 |||||
855 |||||
856 |||||
857 |||||
858 |||||
859 |||||
860 |||||
861 |||||
862 |||||
863 |||||
864 |||||
865 |||||
866 |||||
867 |||||
868 |||||
869 |||||
870 |||||
871 |||||
872 |||||
873 |||||
874 |||||
875 |||||
876 |||||
877 |||||
878 |||||
879 |||||
880 |||||
881 |||||
882 |||||
883 |||||
884 |||||
885 |||||
886 |||||
887 |||||
888 |||||
889 |||||
890 |||||
891 |||||
892 |||||
893 |||||
894 |||||
895 |||||
896 |||||
897 |||||
898 |||||
899 |||||
900 |||||
901 |||||
902 |||||
903 |||||
904 |||||
905 |||||
906 |||||
907 |||||
908 |||||
909 |||||
910 |||||
911 |||||
912 |||||
913 |||||
914 |||||
915 |||||
916 |||||
917 |||||
918 |||||
919 |||||
920 |||||
921 |||||
922 |||||
923 |||||
924 |||||
925 |||||
926 |||||
927 |||||
928 |||||
929 |||||
930 |||||
931 |||||
932 |||||
933 |||||
934 |||||
935 |||||
936 |||||
937 |||||
938 |||||
939 |||||
940 |||||
941 |||||
942 |||||
943 |||||
944 |||||
945 |||||
946 |||||
947 |||||
948 |||||
949 |||||
950 |||||
951 |||||
952 |||||
953 |||||
954 |||||
955 |||||
956 |||||
957 |||||
958 |||||
959 |||||
960 |||||
961 |||||
962 |||||
963 |||||
964 |||||
965 |||||
966 |||||
967 |||||
968 |||||
969 |||||
970 |||||
971 |||||
972 |||||
973 |||||
974 |||||
975 |||||
976 |||||
977 |||||
978 |||||
979 |||||
980 |||||
981 |||||
982 |||||
983 |||||
984 |||||
985 |||||
986 |||||
987 |||||
988 |||||
989 |||||
990 |||||
991 |||||
992 |||||
993 |||||
994 |||||
995 |||||
996 |||||
997 |||||
998 |||||
999 |||||
1000 |||||
1001 |||||
1002 |||||
1003 |||||
1004 |||||
1005 |||||
1006 |||||
1007 |||||
1008 |||||
1009 |||||
1010 |||||
1011 |||||
1012 |||||
1013 |||||
1014 |||||
1015 |||||
1016 |||||
1017 |||||
1018 |||||
1019 |||||
1020 |||||
1021 |||||
1022 |||||
1023 |||||
1024 |||||
1025 |||||
1026 |||||
1027 |||||
1028 |||||
1029 |||||
1030 |||||
1031 |||||
1032 |||||
1033 |||||
1034 |||||
1035 |||||
1036 |||||
1037 |||||
1038 |||||
1039 |||||
1040 |||||
1041 |||||
1042 |||||
1043 |||||
1044 |||||
1045 |||||
1046 |||||
1047 |||||
1048 |||||
1049 |||||
1050 |||||
1051 |||||
1052 |||||
1053 |||||
1054 |||||
1055 |||||
1056 |||||
1057 |||||
1058 |||||
1059 |||||
1060 |||||
1061 |||||
1062 |||||
1063 |||||
1064 |||||
1065 |||||
1066 |||||
1067 |||||
1068 |||||
1069 |||||
1070 |||||
1071 |||||
1072 |||||
1073 |||||
1074 |||||
1075 |||||
1076 |||||
1077 |||||
1078 |||||
1079 |||||
1080 |||||
1081 |||||
1082 |||||
1083 |||||
1084 |||||
1085 |||||
1086 |||||
1087 |||||
1088 |||||
1089 |||||
1090 |||||
1091 |||||
1092 |||||
1093 |||||
1094 |||||
1095 |||||
1096 |||||
1097 |||||
1098 |||||
1099 |||||
1100 |||||
1101 |||||
1102 |||||
1103 |||||
1104 |||||
1105 |||||
1106 |||||
1107 |||||
1108 |||||
1109 |||||
1110 |||||
1111 |||||
1112 |||||
1113 |||||
1114 |||||
1115 |||||
1116 |||||
1117 |||||
1118 |||||
1119 |||||
1120 |||||
1121 |||||
1122 |||||
1123 |||||
1124 |||||
1125 |||||
1126 |||||
1127 |||||
1128 |||||
1129 |||||
1130 |||||
1131 |||||
1132 |||||
1133 |||||
1134 |||||
1135 |||||
1136 |||||
1137 |||||
1138 |||||
1139 |||||
1140 |||||
1141 |||||
1142 |||||
1143 |||||
1144 |||||
1145 |||||
1146 |||||
1147 |||||
1148 |||||
1149 |||||
1150 |||||
1151 |||||
1152 |||||
1153 |||||
1154 |||||
1155 |||||
1156 |||||
1157 |||||
1158 |||||
1159 |||||
1160 |||||
1161 |||||
1162 |||||
1163 |||||
1164 |||||
1165 |||||
1166 |||||
1167 |||||
1168 |||||
1169 |||||
1170 |||||
1171 |||||
1172 |||||
1173 |||||
1174 |||||
1175 |||||
1176 |||||
1177 |||||
1178 |||||
1179 |||||
1180 |||||
1181 |||||
1182 |||||
1183 |||||
1184 |||||
1185 |||||
1186 |||||
1187 |||||
1188 |||||
1189 |||||
1190 |||||
1191 |||||
1192 |||||
1193 |||||
1194 |||||
1195 |||||
1196 |||||
1197 |||||
1198 |||||
1199 |||||
1200 |||||
1201 |||||
1202 |||||
1203 |||||
1204 |||||
1205 |||||
1206 |||||
1207 |||||
1208 |||||
1209 |||||
1210 |||||
1211 |||||
1212 |||||
1213 |||||
1214 |||||
1215 |||||
1216 |||||
1217 |||||
1218 |||||
1219 |||||
1220 |||||
1221 |||||
1222 |||||
1223 |||||
1224 |||||
1225 |||||
1226 |||||
1227 |||||
1228 |||||
1229 |||||
1230 |||||
1231 |||||
1232 |||||
1233 |||||
1234 |||||
1235 |||||
1236 |||||
1237 |||||
1238 |||||
1239 |||||
1240 |||||
1241 |||||
1242 |||||
1243 |||||
1244 |||||
1245 |||||
1246 |||||
1247 |||||
1248 |||||
1249 |||||
1250 |||||
1251 |||||
1252 |||||
1253 |||||
1254 |||||
1255 |||||
1256 |||||
1257 |||||
1258 |||||
1259 |||||
1260 |||||
1261 |||||
1262 |||||
1263 |||||
1264 |||||
1265 |||||
1266 |||||
1267 |||||
1268 |||||
1269 |||||
1270 |||||
1271 |||||
1272 |||||
1273 |||||
1274 |||||
1275 |||||
1276 |||||
1277 |||||
1278 |||||
1279 |||||
1280 |||||
1281 |||||
1282 |||||
1283 |||||
1284 |||||
1285 |||||
1286 |||||
1287 |||||
1288 |||||
1289 |||||
1290 |||||
1291 |||||
1292 |||||
1293 |||||
1294 |||||
1295 |||||
1296 |||||
1297 |||||
1298 |||||
1299 |||||
1300 |||||
1301 |||||
1302 |||||
1303 |||||
1304 |||||
1305 |||||
1306 |||||
1307 |||||
1308 |||||
1309 |||||
1310 |||||
1311 |||||
1312 |||||
1313 |||||
1314 |||||
1315 |||||
1316 |||||
1317 |||||
1318 |||||
1319 |||||
1320 |||||
1321 |||||
1322 |||||
1323 |||||
1324 |||||
1325 |||||
1326 |||||
1327 |||||
1328 |||||
1329 |||||
1330 |||||
1331 |||||
1332 |||||
1333 |||||
1334 |||||
1335 |||||
1336 |||||
1337 |||||
1338 |||||
1339 |||||
1340 |||||
1341 |||||
1342 |||||
1343 |||||
1344 |||||
1345 |||||
1346 |||||
1347 |||||
1348 |||||
1349 |||||
1350 |||||
1351 |||||
1352 |||||
1353 |||||
1354 |||||
1355 |||||
1356 |||||
1357 |||||
1358 |||||
1359 |||||
1360 |||||
1361 |||||
1362 |||||
1363 |||||
1364 |||||
1365 |||||
1366 |||||
1367 |||||
1368 |||||
1369 |||||
1370 |||||
1371 |||||
1372 |||||
1373 |||||
1374 |||||
1375 |||||
1376 |||||
1377 |||||
1378 |||||
1379 |||||
1380 |||||
1381 |||||
1382 |||||
1383 |||||
1384 |||||
1385 |||||
1386 |||||
1387 |||||
1388 |||||
1389 |||||
1390 |||||
1391 |||||
1392 |||||
1393 |||||
1394 |||||
1395 |||||
1396 |||||
1397 |||||
1398 |||||
1399 |||||
1400 |||||
1401 |||||
1402 |||||
1403 |||||
1404 |||||
1405 |||||
1406 |||||
1407 |||||
1408 |||||
1409 |||||
1410 |||||
1411 |||||
1412 |||||
1413 |||||
1414 |||||
1415 |||||
1416 |||||
1417 |||||
1418 |||||
1419 |||||
1420 |||||
1421 |||||
1422 |||||
1423 |||||
1424 |||||
1425 |||||
1426 |||||
1427 |||||
1428 |||||
1429 |||||
1430 |||||
1431 |||||
1432 |||||
1433 |||||
1434 |||||
1435 |||||
1436 |||||
1437 |||||
1438 |||||
1439 |||||
1440 |||||
1441 |||||
1442 |||||
1443 |||||
1444 |||||
1445 |||||
1446 |||||
1447 |||||
1448 |||||
1449 |||||
1450 |||||
1451 |||||
1452 |||||
1453 |||||
1454 |||||
1455 |||||
1456 |||||
1457 |||||
1458 |||||
1459 |||||
1460 |||||
1461 |||||
1462 |||||
1463 |||||
1464 |||||
1465 |||||
1466 |||||
1467 |||||
1468 |||||
1469 |||||
1470 |||||
1471 |||||
1472 |||||
1473 |||||
1474 |||||
1475 |||||
1476 |||||
1477 |||||
1478 |||||
1479 |||||
1480 |||||
1481 |||||
1482 |||||
1483 |||||
1484 |||||
1485 |||||
1486 |||||
1487 |||||
1488 |||||
1489 |||||
1490 |||||
1491 |||||
1492 |||||
1493 |||||
1494 |||||
1495 |||||
1496 |||||
1497 |||||
1498 |||||
1499 |||||
1500 |||||
1501 |||||
1502 |||||
1503 |||||
1504 |||||
1505 |||||
1506 |||||
1507 |||||
1508 |||||
1509 |||||
1510 |||||
1511 |||||
1512 |||||
1513 |||||
1514 |||||
1515 |||||
1516 |||||
1517 |||||
1518 |||||
1519 |||||
1520 |||||
1521 |||||
1522 |||||
1523 |||||
1524 |||||
1525 |||||
1526 |||||
1527 |||||
1528 |||||
1529 |||||
1530 |||||
1531 |||||
1532 |||||
1533 |||||
1534 |||||
1535 |||||
1536 |||||
1537 |||||
1538 |||||
1539 |||||
1540 |||||
1541 |||||
1542 |||||
1543 |||||
1544 |||||
1545 |||||
1546 |||||
1547 |||||
1548 |||||
1549 |||||
1550 |||||
1551 |||||
1552 |||||
1553 |||||
1554 |||||
1555 |||||
1556 |||||
1557 |||||
1558 |||||
1559 |||||
1560 |||||
1561 |||||
1562 |||||
1563 |||||
1564 |||||
1565 |||||
1566 |||||
1567 |||||
1568 |||||
1569 |||||
1570 |||||
1571 |||||
1572 |||||
1573 |||||
1574 |||||
1575 |||||
1576 |||||
1577 |||||
1578 |||||
1579 |||||
1580 |||||
1581 |||||
1582 |||||
1583 |||||
1584 |||||
1585 |||||
1586 |||||
1587 |||||
1588 |||||
1589 |||||
1590 |||||
1591 |||||
1592 |||||
1593 |||||
1594 |||||
1595 |||||
1596 |||||
1597 |||||
1598 |||||
1599 |||||
1600 |||||
1601 |||||
1602 |||||
1603 |||||
1604 |||||
1605 |||||
1606 |||||
1607 |||||
1608 |||||
1609 |||||
1610 |||||
1611 |||||
1612 |||||

```

559	GTCCAGAGCAGATCTCTGCTCCCTCCCTGCTACAGAACTGCGAGCTAAGGCTG	618
204	GTCCAGAGGAGATCTCTGCTCCCTCCCTGCTACAGAACTGCGAGCTAAGGCTG	145
619	CTGGCGAGGCCACCTGTCAATGTCTCTACAGGCAGCCCGTCCCTTCAACCTCATCTTC	678
144	CTGGCGAGGCCACCTGTCAATGTCTCTACAGGCAGCCCGTCCCTTCAACCTCATCTTC	85
679	CAGATATTGCCAGGATGCTGTGTGCTGCTACCCAGAGGCGCGAGGACACCTGCCAG	738
84	CAGATATTGCCAGGATGCTGTGTGCTGCTACCCAGAGGCGCGAGGACACCTGCCAG	25
739	GGTCACTCTGGGGGCCCTGGTC	762
24	GGTGACTCTGGGGGGCCCTGGTC	1
RESULT 3		
LOCUS	BMX436299	
DEFINITION	BMX436299 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YG17	linear EST 15-MAY-2003
ACCESSION	BMX436299	3-PRIME, mRNA sequence.
VERSION	BMX436299.1	GI:30787522
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.	
TITLE	Full-length cDNA libraries and normalization	
JOURNAL	Unpublished (2001)	
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: giref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7995.f For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CS0CAP001AD09FMI&cluster=7995.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Paraday Avenue Genoscope sequence ID : CS0CAP001AD09FMI.	
FEATURES		
source	Location/Qualifiers	
	1. 930	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone="CS0CAP001YG17"	
	/tissue_type="THYMUS"	
	/clone_lib="Homo sapiens THYMUS"	
	/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."	
ORIGIN		
Query Match	12.3%;	Score 135; DB 13; Length 930;
Best Local Similarity	100.0%;	Pred. No. 1.4e-52;
Matches 135;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	811	TTTGCTCTGGACGAGAAACCGCCTGGAGTTTTCACGTGTGGCTACTATGAGCA 870
Db	106	TTTGGCTTGGACGAGAAACCGCCTGGAGTTTTCACGTGTGGCTACTATGAGCA 165
QY	871	TGGATACGGGACGAGTGATGGGTTTCAGAGCTGGGCTGCTTTCCACCCAGCCCGAG 930
Db	166	TGGATACGGGACGAGTGATGGGTTTCAGAGCTGGGCTGCTTTCCACCCAGCCCGAG 225
QY	931	AAGACCCAGTTCAGAT 945

[illegible]

```

Db          218 GTGCCCA 224

RESULT 5
K-EST0101694 S9SNU601 Homo sapiens cDNA clone S9SNU601-51-E07 5',
LOCUS      linear      EST 06-MAR-2002
DEFINITION
BMB28821    318 bp      mRNA
mRNA sequence.
BMB28821    5828821    GI:19185230
BMB28821    5828821    1   GI:19185230
EST.
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 318)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 51 row: E column: 07
High quality sequence stop: 318.
Location/Qualifiers
1. .318
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S9SNU601-51-E07"
/sex="M"
/tissue_type="Ascites"
/cell_type="Epithelial"
/cell_line="GNU-601"
/lab_host="Top10F"
/clone_lib="S9SNU601"
/note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
Site 2: XhoI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
inact mRNA was ligated with DNA-RNA linker including Sfil
site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized with Superscript II using Sfil
oligo-dT primer. After first strand synthesis, RNA was
degraded by NaOH treatment and cDNA was amplified by PCR
reaction. The PCR products were digested with Sfil and
cloned into DraIII- digested pME18S-FL3 vector. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

ORIGIN
Query Match      6.1%; Score 67; DB 12; Length 318;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGCCCTTCTCTGGCCATGCCACAGAGGGGGTCCCTGGGGCCTGGGCAGCTGGGGGCT 60
DB 207 GGGCCCTTCTCTGGGCCATGGCCACAGAGGGGGTCCCTGGGGCCTGGGCAGCTGGGGGCT 266
QY 61 GTGCCCA 67
DB 267 GTGCCCA 273

RESULT 6

```

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 763 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 356.
Location/Qualifiers
1. 377
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1743210"
/dev_stage="19 weeks"
/lab_host="PH108 (ampicillin resistant)"
/clone_lib="Soares fetal lung NBHL19W"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTTACCAATCTGAAGTGGAGCGCCGCAATTTTCTTTT-3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBHL19W."

Query Match 6.1%; Score 67; DB 9; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 67; Conservative 0; Mismatches 0;

QY 1 GGGCCCTTCTCTGGCCATGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 60
|||||
DB 153 GGGCCCTTCTCTGGCCATGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 212
|||||

QY 61 GTGGCCA 67
|||||
DB 213 GTGGCCA 219

Query Match 6.1%; Score 67; DB 9; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 67; Conservative 0; Mismatches 0;

QY 1 GGGCCCTTCTCTGGCCATGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 60
|||||
DB 153 GGGCCCTTCTCTGGCCATGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 212
|||||

QY 61 GTGGCCA 67
|||||
DB 213 GTGGCCA 219

RESULT 9
CB116948 406 bp mRNA linear EST 28-JAN-2003
K-EST0161920 L7N800102 Homo sapiens cDNA clone L7N800102-4-C01 5',
mRNA sequence.
CB116948
CB116948.1 GI:27942755
EST.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 406)
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongseung@kribb.re.kr
Plate: 4 row: C column: 01
High quality sequence stop: 406.
Location/Qualifiers
1. 406

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 763 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 356.
Location/Qualifiers
1. 377
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1743210"
/dev_stage="19 weeks"
/lab_host="PH108 (ampicillin resistant)"
/clone_lib="Soares fetal lung NBHL19W"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTTACCAATCTGAAGTGGAGCGCCGCAATTTTCTTTT-3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBHL19W."

Query Match 6.1%; Score 67; DB 9; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 67; Conservative 0; Mismatches 0;

QY 1 GGGCCCTTCTCTGGCCATGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 60
|||||
DB 153 GGGCCCTTCTCTGGCCATGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 212
|||||

QY 61 GTGGCCA 67
|||||
DB 213 GTGGCCA 219

Query Match 6.1%; Score 67; DB 9; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 67; Conservative 0; Mismatches 0;

QY 1 GGGCCCTTCTCTGGCCATGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 60
|||||
DB 153 GGGCCCTTCTCTGGCCATGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 212
|||||

QY 61 GTGGCCA 67
|||||
DB 213 GTGGCCA 219

RESULT 8
A1193435 377 bp mRNA linear EST 29-OCT-1998
Q558e10.x1 Soares fetal lung NBHL19W Homo sapiens cDNA clone
IMAGE:1743210 3' similar to SW:PSS8_HUMAN Q16651 PROSTASIN
PRECURSOR i, mRNA sequence.
A1193435
A1193435.1 GI:3744644
EST.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 377)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index


```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L7N800102-4-C01"
/sex="M"
/lab_host="Top10P"
/clone_lib="L7N800102"
/notes="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI; Site 2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dT-selected mRNA by priming with dT-tailed vector. The dT-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10P by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

ORIGIN
Query Match 6.1%; Score 67; DB 14; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 67; Conservative 0; Mismatches 0;

QY 1 GGGCCCTTCTCTGGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGGAGTGGGGGCT 60
Db 209 GGGCCCTTCTCTGGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGGAGTGGGGGCT 268

QY 61 GTGGCCA 67
Db 269 GTGGCCA 275

RESULT 10
LOCUS AU298208
DEFINITION AU298208 male adult skin, full-length enriched chimpanzee cDNA library. Pan troglodytes versus cDNA clone PstA2814 5' similar to human RefSeq mRNA NM_002773, mRNA sequence.
ACCESSION AU298208
VERSION AU298208.1 GI:29532507
KEYWORDS EST.
SOURCE Pan troglodytes versus
ORGANISM Pan troglodytes versus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (bases 1 to 409)
Sakate,R., Osada,N., Hida,M., Sugano,S., Hayaaka,I., Shimohira,N., Yanagi,S., Suto,Y., Hashimoto,K. and Hirai,M.
Analysis of 5'-end sequences of chimpanzee cDNAs
Genome Res. 13 (5), 1022-1026 (2003)
22612989
12727913
Contact: Momoki Hirai
Department of Integrated Biosciences
The University of Tokyo, Graduate School of Frontier Sciences
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan
Tel: 81-4-7136-3688
Fax: 81-4-7136-3687
Email: mhirai@k.u-tokyo.ac.jp.
Location/Qualifiers
1. 409
/organism="Pan troglodytes versus"
/mol_type="mRNA"
/sub_species="verus"
/db_xref="taxon:37012"
/clone="PstA2814"
/sex="male"

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L7N800102-4-C01"
/sex="M"
/lab_host="Top10P"
/clone_lib="L7N800102"
/notes="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI; Site 2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dT-selected mRNA by priming with dT-tailed vector. The dT-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10P by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

ORIGIN
Query Match 6.1%; Score 67; DB 9; Length 409;
Best Local Similarity 100.0%; Pred. No. 2e-20; Indels 0; Gaps 0;
Matches 67; Conservative 0; Mismatches 0;

QY 1 GGGCCCTTCTCTGGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGGAGTGGGGGCT 60
Db 213 GGGCCCTTCTCTGGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGGAGTGGGGGCT 272

QY 61 GTGGCCA 67
Db 273 GTGGCCA 279

RESULT 11
LOCUS AI343968
DEFINITION tc01d05.x1 NCI CGAP Col6 Homo sapiens cDNA clone IMAGE:2062569 3', similar to TR:Q16651 Q16651 PROTSIN. ;, mRNA sequence.
ACCESSION AI343968
VERSION AI343968.1 GI:4081174
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 462)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 586 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.
Location/Qualifiers
1. 462
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2062569"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/clone_lib="NCI CGAP Col6"
/notes="Organ: colon; Vector: pT7M3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; plasmid DNA from the normalized library NCI CGAP Col6 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1057416-1061255, and 1144584-1145351). Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN
Query Match 6.1%; Score 67; DB 9; Length 462;
Best Local Similarity 100.0%; Pred. No. 2e-20; Indels 0; Gaps 0;
Matches 67; Conservative 0; Mismatches 0;

FEATURES
source

```

```

QY 1 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 60
Db 127 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 186
QY 61 GTGGCCA 67
Db 187 GTGGCCA 193

RESULT 12
AI761519 472 bp mRNA linear EST 21-DEC-1999
LOCUS similar02.xl NCI CGAP Col6 Homo sapiens cDNA clone IMAGE:2394723 3'
DEFINITION similar to SW:FS98_HUMAN Q16651 PROSTASIN PRECURSOR ;, mRNA
AUTHORS sequence.
TITLE AI761519
ACCESSION AI761519.1 GI:5177186
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 472)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 601 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 451.
Location/Qualifiers
FEATURES
source
1..472
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2394723"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/clone_lib="NCI-CGAP Col6"
/note="Organ: colon; Vector: pVT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP Col6 was
prepared, and as circles were made in vitro. Following HAP
hybridization, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "

Query Match 6.1%; Score 67; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 2e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 60
Db 102 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 161
QY 61 GTGGCCA 67
Db 162 GTGGCCA 168

RESULT 13
AI761519 479 bp mRNA linear EST 08-MAY-2003
LOCUS similar02.xl NCI CGAP Col6 Homo sapiens cDNA clone PstA5634 5' similar to
DEFINITION human RefSeq mRNA NM_002773, mRNA sequence.
AUTHORS sequence.
TITLE AI761519
ACCESSION AI761519.1 GI:29532616
VERSION EST.
KEYWORDS Pan troglodytes verus
SOURCE Pan troglodytes verus
ORGANISM Pan troglodytes verus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 479)
AUTHORS Sakate, R., Osada, N., Hida, M., Sugano, S., Hayasaka, I., Shimohira, N.,
Yanagi, S., Suto, Y., Hashimoto, K. and Hirai, M.
Analysis of 5'-end sequences of chimpanzee cDNAs
Genome Res. 13 (5), 1022-1026 (2003)
JOURNAL 22612989
MEDLINE 12727913
PUBMED
COMMENT Contact: Momoki Hirai
Department of Integrated Biosciences
The University of Tokyo, Graduate School of Frontier Sciences
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan
Tel: 81-4-7136-3688
Fax: 81-4-7136-3687
Email: mhiraik@k.u-tokyo.ac.jp.
Location/Qualifiers
FEATURES
source
1..479
/organism="Pan troglodytes verus"
/mol_type="mRNA"
/sub_species="verus"
/db_xref="taxon:37012"
/clone="PstA5634"
/sex="male"
/tissue_type="skin"
/dev_stage="adult"
/clone_lib="male adult skin, full-length enriched
chimpanzee cDNA library"

Query Match 6.1%; Score 67; DB 9; Length 479;
Best Local Similarity 100.0%; Pred. No. 2e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 60
Db 212 GGGCCCTTGTCTGGGCGCATGGCCAGAGGGGTCTCTGGGGCTGGGCGAGCTGGGGCT 271
QY 61 GTGGCCA 67
Db 272 GTGGCCA 278

RESULT 14
CB996389 479 bp mRNA linear EST 01-MAY-2003
LOCUS AGENCOURT_13621004 NIH_MGC_148 Homo sapiens cDNA clone
DEFINITION IMAGE:30331484 5', mRNA sequence.
AUTHORS sequence.
TITLE CB996389
ACCESSION CB996389.1 GI:30290909
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 479)
AUTHORS NIH-MGC http://mgc.ncbi.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Dr. Stefan Hansson

```

cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help and advice from Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAW346 row: P column: 21
High quality sequence stop: 478.

FEATURES

source

1. .479
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30331484"
/tissue_type="pre-eclamptic placenta"
/lab_host="DH10B TonA"
/clone_lib="NIH MGC 148"
/note="Organ: placenta; Vector: pBluescriptR; Site 1: all-XhoI; Site 2: BamH; Library is oligo-dT primed and directionally cloned using primer
5'-TTTTTTTTTTTTTTTNN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 6.1%; Score 67; DB 14; Length 479;
Best Local Similarity 100.0%; Pred. No. 2e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGCCCTTGTCTCGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGCTGGGGCT 60
Db 239 GGCCCTTGTCTCGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGCTGGGGCT 298
QY 61 GTGGCCA 67
Db 299 GTGGCCA 305

RESULT 15

BF063417
LOCUS 7h89e10.x1 NCI CGAP Col6 Homo sapiens cDNA clone IMAGE:3323178 3'
DEFINITION similar to SW:FS88_HUMAN Q16651 PROSTASIN PRECURSOR ;, mRNA
sequence.
ACCESSION BF063417 GI:10822327
VERSION BF063417.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 492)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contract: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco.
Location/Qualifiers

FEATURES

source

1. .492
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3323178"
/tissue_type="colon tumor, RER+"
/lab_host="DH108"
/clone_lib="NCI CGAP Col6"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; plasmid DNA from the normalized library NCI CGAP Col6 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1057416-1061255, and 1144584-1145351). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 6.1%; Score 67; DB 10; Length 492;
Best Local Similarity 100.0%; Pred. No. 2e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGCCCTTGTCTCGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGCTGGGGCT 60
Db 153 GGCCCTTGTCTCGGCCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGCTGGGGCT 212
QY 61 GTGGCCA 67
Db 213 GTGGCCA 219

Search completed: February 25, 2004, 14:35:31
Job time : 2787 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 12:20:46 ; Search time 110 Seconds
(without alignments)
5559.599 Million cell updates/sec

Title: US-10-037-417-45
Perfect score: 1102
Sequence: 1 gggccctgtctgggcaat.....ggggttcgtatggggcctcc 1102

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : Issued Patents NA:*
- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
 - 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
 - 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
 - 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
 - 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
 - 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	24	2.2	610	4	US-09-280-116-22	Sequence 22, Appl
2	23	2.1	340	4	US-09-280-116-28	Sequence 28, Appl
C 3	23	2.1	340	4	US-09-280-116-29	Sequence 29, Appl
C 4	22	2.0	1142	4	US-09-386-642-8	Sequence 8, Appl
5	22	2.0	1169	4	US-09-386-642-7	Sequence 7, Appl
C 6	21	1.9	90	3	US-09-020-956-120	Sequence 120, App
C 7	21	1.9	90	3	US-09-030-607-120	Sequence 120, App
C 8	21	1.9	90	4	US-09-439-313-120	Sequence 120, App
C 9	21	1.9	90	4	US-09-352-616A-120	Sequence 120, App
C 10	21	1.9	90	4	US-09-232-149A-120	Sequence 120, App
C 11	21	1.9	90	4	US-09-159-812-120	Sequence 120, App
C 12	21	1.9	90	4	US-09-636-215-120	Sequence 120, App
C 13	21	1.9	90	4	US-09-685-166A-120	Sequence 120, App
C 14	21	1.9	90	4	US-09-115-453-120	Sequence 120, App
C 15	21	1.9	90	4	US-09-688-489-120	Sequence 120, App
16	21	1.9	504	4	US-09-636-215-839	Sequence 839, App
17	21	1.9	504	4	US-09-685-166A-839	Sequence 839, App
18	21	1.9	700	4	US-09-280-116-68	Sequence 68, Appl
19	21	1.9	765	4	US-09-439-313-524	Sequence 524, App
20	21	1.9	765	4	US-09-636-215-524	Sequence 524, App
21	21	1.9	765	4	US-09-685-166A-524	Sequence 524, App
22	21	1.9	871	1	US-08-744-026-2	Sequence 2, Appl
23	21	1.9	871	2	US-09-102-732-2	Sequence 2, Appl
24	21	1.9	871	3	US-09-261-767-2	Sequence 2, Appl
25	21	1.9	871	4	US-08-969-987-7	Sequence 7, Appl
26	21	1.9	1037	4	US-09-386-642-60	Sequence 60, Appl
27	21	1.9	1167	3	US-09-020-956-175	Sequence 175, App

ALIGNMENTS

RESULT 1

US-09-280-116-22
; Sequence 22, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 610
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: trypsin-like serine proteases
; NAME/KEY: misc feature
; LOCATION: (1)..(610)
; OTHER INFORMATION: n = a, t, c, or g
US-09-280-116-22

Query Match 2.2%; Score 24; DB 4; Length 610;
Best Local Similarity 100.0%; Pred. No. 0.082;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 TGGCCTTGGCAAGTGAGCCTGCAC 213
Db 102 TGGCCTTGGCAAGTGAGCCTGCAC 125

RESULT 2

US-09-280-116-28/c
; Sequence 28, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-28
Query Match 2.1%; Score 23; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 745 TCTGGGGGCCCCCTGCTGTGA 767
Db 241 TCTGGGGGCCCCCTGCTGTGA 219
RESULT 3
US-09-280-116-29/c
Sequence 29, Application US/09280116A
Patent No. 6331427
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
FILE REFERENCE: 5800-24, 035800/176965
CURRENT APPLICATION NUMBER: US/09/280,116A
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 268
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 29
LENGTH: 340
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-29
Query Match 2.1%; Score 23; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 745 TCTGGGGGCCCCCTGCTGTGA 767
Db 241 TCTGGGGGCCCCCTGCTGTGA 219
RESULT 4
US-09-386-642-8
Sequence 8, Application US/09386642
Patent No. 6420157
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jenson
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1142
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-8
Query Match 2.0%; Score 22; DB 4; Length 1142;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 731 CCTGCCAGGGTGACTCTGGGG 752
Db 704 CCTGCCAGGGTGACTCTGGGG 725

RESULT 5
US-09-386-642-7
Sequence 7, Application US/09386642
Patent No. 6420157
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jenson
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1169
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-7
Query Match 2.0%; Score 22; DB 4; Length 1169;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 731 CCTGCCAGGGTGACTCTGGGG 752
Db 731 CCTGCCAGGGTGACTCTGGGG 752
RESULT 6
US-09-020-956-120/c
Sequence 120, Application US/09020956
Patent No. 6261562
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillin, Davin C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
NUMBER OF SEQUENCES: 178
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,956
FILING DATE: 09-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

```
US-09-020-956-120
Query Match          1.9%; Score 21; DB 3; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 739 GGTGACTCTGGGGGCCCTG 759
    |||||
Db 34 GGTGACTCTGGGGGCCCTG 14

RESULT 7
US-09-030-607-120/c
; Sequence 120, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 120:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-030-607-120

Query Match          1.9%; Score 21; DB 3; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 739 GGTGACTCTGGGGGCCCTG 759
    |||||
Db 34 GGTGACTCTGGGGGCCCTG 14

RESULT 8
US-09-439-313-120/c
; Sequence 120, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui

US-09-020-956-120
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(90)
; OTHER INFORMATION: n = A,T,C or G
US-09-439-313-120

Query Match          1.9%; Score 21; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 739 GGTGACTCTGGGGGCCCTG 759
    |||||
Db 34 GGTGACTCTGGGGGCCCTG 14

RESULT 9
US-09-352-616A-120/c
; Sequence 120, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqui
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(90)
; OTHER INFORMATION: n = A,T,C or G
US-09-352-616A-120

Query Match          1.9%; Score 21; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 739 GGTGACTCTGGGGGCCCTG 759
    |||||
Db 34 GGTGACTCTGGGGGCCCTG 14

RESULT 10
US-09-232-149A-120/c
; Sequence 120, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui
```

```

; APPLICANT: Xu, Jiangchun.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232,149A
; CURRENT FILING DATE: 1998-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(90)
; OTHER INFORMATION: n = A,T,C or G
US-09-232-149A-120

Query Match      1.9%; Score 21; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 GGTGACTCTGGGGGCCCTG 759
DB 34 GGTGACTCTGGGGGCCCTG 14

RESULT 11
US-09-159-812-120/c
; Sequence 120, Application US/09159812A
; Patent No. 6613872
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF
; TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.428C5
; CURRENT APPLICATION NUMBER: US/09/159,812A
; CURRENT FILING DATE: 1998-09-23
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(90)
; OTHER INFORMATION: n = A,T,C or G
US-09-159-812-120

Query Match      1.9%; Score 21; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 GGTGACTCTGGGGGCCCTG 759
DB 34 GGTGACTCTGGGGGCCCTG 14

RESULT 12
US-09-636-215-120/c
; Sequence 120, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.

```

```

; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(90)
; OTHER INFORMATION: n = A,T,C or G
US-09-636-215-120

Query Match      1.9%; Score 21; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 GGTGACTCTGGGGGCCCTG 759
DB 34 GGTGACTCTGGGGGCCCTG 14

RESULT 13
US-09-685-166A-120/c
; Sequence 120, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(90)

```


Thu Feb 26 13:50:21 2004

```

; OTHER INFORMATION: n = A,T,C or G
US-09-685-166A-120
Query Match 1.9%; Score 21; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 GGTGACTCTGGGGGCCCTG 759
Db 34 GGTGACTCTGGGGGCCCTG 14

Search completed: February 25, 2004, 14:37:33
Job time : 111 secs

```

```

RESULT 14
US-09-115-453-120/c
; Sequence 120, Application US/09115453B
; Patent No. 6657056
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND
; TITLE OF INVENTION: METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427C4
; CURRENT APPLICATION NUMBER: US/09/115,453B
; CURRENT FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(90)
; OTHER INFORMATION: n = A,T,C or G
US-09-115-453-120

```

```

Query Match 1.9%; Score 21; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 739 GGTGACTCTGGGGGCCCTG 759
Db 34 GGTGACTCTGGGGGCCCTG 14

```

```

RESULT 15
US-09-688-489-120/c
; Sequence 120, Application US/09688489
; Patent No. 6664377
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427D2
; CURRENT APPLICATION NUMBER: US/09/688,489
; CURRENT FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(90)
; OTHER INFORMATION: n = A,T,C or G
US-09-688-489-120

```

```

Query Match 1.9%; Score 21; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2004, 13:49:01 ; Search time 454 Seconds

(without alignments)
8499.586 Million cell updates/sec

Title: US-10-037-417-45

Perfect score: 1102

Sequence: 1 gggcccttgctgggcca.....gggtcttgatgggctcc 1102

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2308684 seqs, 1750822206 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*

17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	716	65.0	2457	9	US-09-888-615-52
2	135	12.3	768	14	US-10-221-097-10
c 3	67	6.1	315	14	Sequence 21306, A
c 4	67	6.1	543	14	Sequence 7606, Ap
5	67	6.1	1733	14	Sequence 85, Appl
6	67	6.1	3382	14	Sequence 447, Ap
7	53	4.8	1726	12	US-10-101-510-447
8	53	4.8	1834	9	US-10-042-865-29
9	53	4.8	1834	9	US-09-948-094-1
10	53	4.8	1834	9	US-09-880-107-2214
11	53	4.8	1834	9	US-09-967-768A-141
c 12	51	4.6	596	9	Sequence 141, Ap
c 13	51	4.6	596	9	Sequence 261, Ap
c 14	51	4.6	596	9	US-09-922-217-931
15	49	4.4	1161	12	US-09-833-263-931
					Sequence 931, Appl
					Sequence 931, Appl
					Sequence 31, Appl

16	42	3.8	1668	9	US-09-925-301-208	Sequence 208, Appl
c 17	37	3.4	1005	9	US-09-764-864-710	Sequence 710, Appl
c 18	37	3.4	1282	9	US-09-764-864-288	Sequence 288, Appl
c 19	25	2.3	506	9	US-09-783-590-5961	Sequence 5961, Ap
20	24	2.2	402	9	US-09-879-792-1	Sequence 1, Appli
21	23	2.1	1138	9	US-09-888-615-58	Sequence 58, Appl
22	23	2.1	1169	9	US-09-804-156-10	Sequence 10, Appl
23	23	2.1	1169	13	US-10-067-761-10	Sequence 10, Appl
24	23	2.1	1169	14	US-10-319-519-10	Sequence 1, Appli
25	23	2.1	1683	9	US-09-912-559-1	Sequence 2, Appli
26	23	2.1	1683	9	US-09-912-559-2	Sequence 1, Appli
27	23	2.1	1683	15	US-10-391-215-1	Sequence 2, Appli
28	23	2.1	1683	15	US-10-391-215-2	Sequence 3, Appli
29	23	2.1	1683	15	US-10-391-215-3	Sequence 4, Appli
30	23	2.1	1683	15	US-10-391-215-4	Sequence 1668, Ap
31	23	2.1	3008	9	US-09-880-107-1668	Sequence 81, Appl
32	22	2.0	174	14	US-10-131-409-81	Sequence 81, Appl
33	22	2.0	174	15	US-10-150-811-81	Sequence 1, Appli
34	22	2.0	296	14	US-10-240-730-1	Sequence 19, Appl
35	22	2.0	870	14	US-10-131-409-19	Sequence 33, Appl
36	22	2.0	870	15	US-10-150-811-19	Sequence 33, Appl
37	22	2.0	882	12	US-10-042-865-33	Sequence 34, Appl
38	22	2.0	882	12	US-10-042-865-34	Sequence 3, Appli
39	22	2.0	1113	14	US-10-240-730-3	Sequence 68, Appl
40	22	2.0	1375	9	US-09-764-898-68	Sequence 1, Appli
41	22	2.0	1797	14	US-10-103-616-1	Sequence 29, Appl
42	22	2.0	2255	12	US-10-399-645-29	Sequence 19, Appl
43	22	2.0	3316	15	US-10-156-214A-19	Sequence 38, Appl
44	22	2.0	3387	9	US-09-888-615-38	Sequence 23, Appl
45	22	2.0	3711	12	US-10-399-645-23	

ALIGNMENTS

RESULT 1

US-09-888-615-52

Sequence 52, Application US/09888615

Patent No. US20020064856A1

GENERAL INFORMATION:

APPLICANT: PLOWMAN, GREGORY

APPLICANT: WHYTE, DAVID

APPLICANT: CHAREPPEL, SEAN

APPLICANT: CHARVDCZAK, GLEN

APPLICANT: MANNING, GERARD

APPLICANT: SUDARSANAM, SUCHA

TITLE OF INVENTION: NOVEL PROTEASES

FILE REFERENCE: 038602/1214

CURRENT APPLICATION NUMBER: US/09/888, 615

CURRENT FILING DATE: 2001-06-26

PRIOR APPLICATION NUMBER: 60/214, 047

PRIOR FILING DATE: 2000-06-26

NUMBER OF SEQ ID NOS: 150

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 52

LENGTH: 2457

TYPE: DNA

ORGANISM: Homo sapiens

US-09-888-615-52

Query Match	65.0%	Score 716;	DB 9;	Length 2457;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 816;	Conservative	0;	Mismatches	2;
			Indels	0;
			Gaps	0;
Qy	128	ACTGCGGGCGCCCTGAGCCCTCGCGCGGCTCGTGGGGGGCTCAAAACGCGAGCGCGGCA	187	
Db	110	ACTGCGGGCGCCCTGAGCCCTCGCGCGGCTCGTGGGGGGCTCAAAACGCGAGCGCGGCA	169	
Qy	188	CTTGGCTTGGCAAGTAGGCTGCACCATGGAGTGGCCACATCTGCGGGGGCTCCCTCA	247	
Db	170	CTTGGCTTGGCAAGTAGGCTGCACCATGGAGTGGCCACATCTGCGGGGGCTCCCTCA	229	
Qy	248	TCGCCCCCTCTGGGTCTCTCCGCTGCTCACTGTTTCATGACGATGGGACGTTGGAGC	307	

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 768
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-221-097-10

Query Match      12.3%; Score 135; DB 14; Length 768;
Best Local Similarity 100.0%; Pred. No. 6e-59;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 ATCTGGGGGCTCAAAAGCGCGACGCGGCGCTTGGCAAGTGAGCTGCACCAT 216
DB 1 ATCTGGGGGCTCAAAAGCGCGACGCGGCGCTTGGCAAGTGAGCTGCACCAT 60

QY 217 GGAGGTGGCCACATCTGCGGGGGCTCCCTCATCGCCCCCTCTGGGTCTCTCCGCTGCT 276
DB 61 GGAGGTGGCCACATCTGCGGGGGCTCCCTCATCGCCCCCTCTGGGTCTCTCCGCTGCT 120

QY 277 CACTGTTTCATGACG 291
DB 121 CACTGTTTCATGACG 135

RESULT 3
US-10-029-386-21306/c
; Sequence 21306, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21306
; LENGTH: 315
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
; OTHER INFORMATION: NT HIT: g11479472, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q1651, EVALUE 1.00e-08
; OTHER INFORMATION: EST_HUMAN HIT: AU142128.1, EVALUE 0.00e+00
US-10-029-386-21306

Query Match      6.1%; Score 67; DB 14; Length 315;
Best Local Similarity 100.0%; Pred. No. 5.5e-24;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTCTCTGGGCGCATGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 60
DB 104 GGGCCCTTCTCTGGGCGCATGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCT 45

QY 61 GTGGCCA 67
DB 44 GTGGCCA 38

RESULT 4
US-10-029-386-7606/c
; Sequence 7606, Application US/10029386
; Publication No. US20030194704A1
; NUMBER OF SEQ ID NOS: 52

```

```

; APPLICANT: Agawal, Pankaj
; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoxing
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50016
; CURRENT APPLICATION NUMBER: US/10/221,097
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: PCT/US01/07143
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/187,107
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: 60/236,874
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/188,916
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/237,846
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 52

```

```

; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7606
; LENGTH: 543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
; OTHER INFORMATION: SWISSPROT HIT: Q16651, EVALUE 7.00e-04
; OTHER INFORMATION: EST HUMAN HIT: AUI42128.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: U33446.1, EVALUE 0.00e-00
; US-10-029-386-7606

Query Match      6.1%; Score 67; DB 14; Length 543;
Best Local Similarity 100.0%; Pred. No. 5.1e-24;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTGGGGCTGGGCGCTGGGGGCT 60
DB 227 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTGGGGCTGGGCGCTGGGGGCT 168

QY 61 GTGGCCA 67
DB 167 GTGGCCA 161

RESULT 5
US-10-176-847-85
; Sequence 85, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Velby, Pette Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MFI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 85
; LENGTH: 1733
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-176-847-85

Query Match      6.1%; Score 67; DB 14; Length 1733;
Best Local Similarity 100.0%; Pred. No. 4.4e-24;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTGGGGCTGGGCGCTGGGGGCT 60
DB 81 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTGGGGCTGGGCGCTGGGGGCT 140

QY 61 GTGGCCA 67
DB 141 GTGGCCA 147
```

```

RESULT 6
US-10-101-510-447
; Sequence 447, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 447
; LENGTH: 3382
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-101-510-447

Query Match      6.1%; Score 67; DB 14; Length 3382;
Best Local Similarity 100.0%; Pred. No. 4e-24;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTGGGGCTGGGCGCTGGGGGCT 60
DB 212 GGGCCCTTGTCTGGCCATGGCCAGAGGGGCTCTGGGGCTGGGCGCTGGGGGCT 271

QY 61 GTGGCCA 67
DB 272 GTGGCCA 278

RESULT 7
US-10-042-865-29
; Sequence 29, Application US/10042865
; Publication No. US20040029216A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Casman, Stacie J
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zhong, Mei
; APPLICANT: Gangolli, Baha A
; APPLICANT: Burgess, Catherine E
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine A.M
; APPLICANT: Taylor, Sarah
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Miller, Charles E
; APPLICANT: Guo, Xiaojia
; APPLICANT: Boldog, Ference L
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Gerlach, Valerie L
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Rothenberg, Mark E
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John
; APPLICANT: Smithson, Glennda
; APPLICANT: Gunther, Erik
; APPLICANT: Stone, David
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; TITLE OF INVENTION: Using the Same
; FILE REFERENCE: 21402-537
; CURRENT APPLICATION NUMBER: US/10/042,865
```

; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/260,417
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/260,831
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/272,338
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/274,876
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/284,704
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 264
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 1726
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-865-29

Query Match 4.8%; Score 53; DB 12; Length 1726;
Best Local Similarity 100.0%; Pred. No. 6.8e-17; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0;

QY 15 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCA 67
Db 225 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCA 277

RESULT 8
US-09-948-094-1
; Sequence 1, Application US/09948094
; Patent No. US20020090625A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Mok, Samuel
; TITLE OF INVENTION: Methods of Detecting Cancer Based on Prostatein

; FILE REFERENCE: 81994/282423
; CURRENT APPLICATION NUMBER: US/09/948,094
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (229)..(1260)
US-09-948-094-1

Query Match 4.8%; Score 53; DB 9; Length 1834;
Best Local Similarity 100.0%; Pred. No. 6.7e-17; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0;

QY 15 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCA 67
Db 225 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCA 277

RESULT 9
US-09-980-107-2214
; Sequence 2214, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/980,107
; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2214
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L41351
US-09-880-107-2214

Query Match 4.8%; Score 53; DB 9; Length 1834;
Best Local Similarity 100.0%; Pred. No. 6.7e-17; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0;

QY 15 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCA 67
Db 225 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCA 277

RESULT 10
US-09-967-768A-141
; Sequence 141, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena

; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 141
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-141

Query Match 4.8%; Score 53; DB 9; Length 1834;
Best Local Similarity 100.0%; Pred. No. 6.7e-17; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0;

QY 15 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCA 67
Db 225 GGCCATGGCCAGAGGGGGTCTCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCA 277

RESULT 11
US-10-097-340-261
; Sequence 261, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS

APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumei ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
FILE REFERENCE: Assessment, Prevention, and Therapy of Ovarian Cancer
FILE REFERENCE: MRI-030
CURRENT FILING DATE: 2002-03-14
CURRENT APPLICATION NUMBER: US/10/097,340
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 261
LENGTH: 1834
TYPE: DNA
ORGANISM: Homo sapiens
US-10-097-340-261

Query Match 4.8%; Score 53; DB 14; Length 1834;
Best Local Similarity 100.0%; Pred. No. 6.7e-17; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 0

Qy 15 GGCCATGCCCCAGAGGGGGTCTGGGGCTGGCCAGCTGGGGCTGTGGCCA 67
Db 225 GGCCATGCCCCAGAGGGGGTCTGGGGCTGGCCAGCTGGGGCTGTGGCCA 277

RESULT 12
US-09-922-217-931/c
Sequence 931, Application US/09922217
Publication No. US200207641A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yuqiu
APPLICANT: Smith, Carole Lynn
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C3
CURRENT APPLICATION NUMBER: US/09/922,217
CURRENT FILING DATE: 2001-08-03
NUMBER OF SEQ ID NOS: 1124
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 931
LENGTH: 596
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 538_
OTHER INFORMATION: n = A,T,C or G

US-09-922-217-931

Query Match 4.6%; Score 51; DB 9; Length 596;
Best Local Similarity 100.0%; Pred. No. 8.3e-16; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 0

Qy 1 GGCCCTTGTCTGGCCATGGCCCAAGAGGGGTCTGGGGCTGGGCAG 51
Db 589 GGCCCTTGTCTGGCCATGGCCCAAGAGGGGTCTGGGGCTGGGCAG 539

RESULT 13
US-09-833-263-931/c
Sequence 931, Application US/09833263
Patent No. US20020110547A1
GENERAL INFORMATION:
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Stolk, John A.
APPLICANT: Meagher, Madeleine J.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
FILE REFERENCE: 210121.471C12
CURRENT APPLICATION NUMBER: US/09/833,263
CURRENT FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 1093
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 931
LENGTH: 596
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)-(596)
OTHER INFORMATION: n = A,T,C or G
US-09-833-263-931

Query Match 4.6%; Score 51; DB 9; Length 596;
Best Local Similarity 100.0%; Pred. No. 8.3e-16; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 0

Qy 1 GGCCCTTGTCTGGCCATGGCCCAAGAGGGGTCTGGGGCTGGGCAG 51
Db 589 GGCCCTTGTCTGGCCATGGCCCAAGAGGGGTCTGGGGCTGGGCAG 539

RESULT 14
US-10-025-380-931/c
Sequence 931, Application US/10025380
Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yuqiu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Darrick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025,380
CURRENT FILING DATE: 2001-12-19
NUMBER OF SEQ ID NOS: 1129

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 931
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 538
; OTHER INFORMATION: n = A,T,C or G
US-10-025-380-931

Query Match      4.6%; Score 51; DB 13; Length 596;
Best Local Similarity 100.0%; Pred. No. 8.3e-16;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCTTCTCTGGGCGCATGGCCAGAGAGGGGGTCTCTGGGGCCTGGGCAG 51
Db 589 GGGCCCTTCTCTGGGCGCATGGCCAGAGAGGGGGTCTCTGGGGCCTGGGCAG 539

RESULT 15
US-10-042-865-31
; Sequence 31, Application US/10042865
; Publication No. US20040029216A1
; GENERAL INFORMATION:
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Casman, Stacie J
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zhong, Mei
; APPLICANT: Gangolli, Esha A
; APPLICANT: Burgess, Catherine E
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine A.M
; APPLICANT: Taylor, Sarah
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Miller, Charles E
; APPLICANT: Guo, Xiaojia
; APPLICANT: Boldog, Ference L
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Gerlach, Valerie L
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Rothenberg, Mark E
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John
; APPLICANT: Smithson, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Stone, David
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-537
; CURRENT APPLICATION NUMBER: US/10/042,865
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/260,417
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/260,831
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/272,338
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/274,876
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/284,704
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 264
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 1161
```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-865-31

Query Match      4.4%; Score 49; DB 12; Length 1161;
Best Local Similarity 100.0%; Pred. No. 8.1e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 ATGGCCCAAGAGGGGGTCTCTGGGGCCTGGGCAGCTGGGGGCTGTGGCCA 67
Db 1 ATGGCCCAAGAGGGGGTCTCTGGGGCCTGGGCAGCTGGGGGCTGTGGCCA 49

Search completed: February 25, 2004, 15:44:57
Job time : 457 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 14:35:42 ; Search time 100 Seconds
(without alignments)
1008.695 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953
Sequence: 1 MAQKGVLPGLGAVANS...TKSLVLPWLSPLSLGLMGF 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1953	100.0	357	5	ABB09524 Human pro
2	1874	96.0	344	5	ABB09523 Human pro
3	1505	77.1	818	5	ABB98135 Human PMM
4	1505	77.1	818	5	AAU82753 Amino aci
5	1202	61.5	766	4	AAU93568 Human pro
6	970	49.7	178	5	AAU75082 Human pro
7	804.5	41.2	198	4	ABG04853 Novel hum
8	690	35.3	343	5	ABG96402 Human ova
9	690	35.3	343	5	AAU78547 Human pro
10	690	35.3	343	5	ABG07285 Amino aci
11	690	35.3	343	6	ABU37067 Human bre
12	690	35.3	343	7	ADD47565 Human pro
13	690	35.3	343	7	ADD47561 Human pro
14	645	33.0	307	5	ABG98415 Human NOV
15	627.5	32.1	339	6	ABG72018 Mouse cha
16	612	31.3	386	5	ABG98416 Human NOV
17	593.5	30.4	282	5	ABP61011 Novel hum
18	590	30.2	272	5	ABG07286 Human pro
19	590	30.2	280	5	ABP61010 Novel hum
20	590	30.2	280	5	ABG91414 Primate L
21	590	30.2	284	5	AAU00467 Human ser
22	582.5	29.8	346	5	AAU74748 Human pro
23	582	29.8	691	5	AAE14348 Human pro
24	575.5	29.5	255	4	AA667514 Amino aci
25	574	29.4	262	5	ABB07284 Human pro

26	568.5	29.1	328	4	AAU41174	Aam41174 Human pol
27	567	29.0	389	5	AAU75907	Aau75907 Human epi
28	566	29.0	297	2	AAW73304	Aaw77304 Amino aci
29	563	28.8	290	3	AAW73388	Aay73388 HTRM clon
30	563	28.8	290	4	AAU12282	Aau12282 Human pro
31	563	28.8	290	4	AAU12282	Aab73945 Human pro
32	563	28.8	290	4	AAU12282	Aae03821 Human gen
33	563	28.8	290	5	ABG64545	Abg64545 Human alb
34	563	28.8	290	6	ABO17726	Abol7726 Novel hum
35	563	28.8	290	6	ABU80980	Abu80980 Human pro
36	563	28.8	290	6	ABU66680	Abu66680 Human pro
37	563	28.8	290	6	ABU59761	Abu59761 Novel sec
38	563	28.8	290	6	ABO24951	Abu24951 Human sec
39	563	28.8	290	6	ABG73394	Abg73394 Human ser
40	563	28.8	290	6	ABU66956	Abu66956 Human sec
41	563	28.8	290	6	ADA45741	Ada45741 Novel hum
42	563	28.8	290	6	ADA76172	Ada76172 Human pro
43	563	28.8	290	6	ADA18822	Ada18822 Human pro
44	563	28.8	290	6	ADA61445	Ada61445 Homo sapi
45	563	28.8	290	6	ADB19230	Adb19230 Novel hum

ALIGNMENTS

RESULT 1

ABB09524
ID ABB09524 standard; protein; 357 AA.

XX ABB09524;

XX 01-NOV-2002 (first entry)

DE Human prostatic precursor-like NOV14b protein, SEQ ID NO:46.

Human; NOVX; neurological disorder; Alzheimer's disease;
Huntington's disease; Parkinson's disease; pain; behavioural disorder;
addiction; tubercous sclerosis; cancer; immune disorder; allergy;
autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;
thyroiditis; cardiovascular disease; hypertension; reproductive disorder;
endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;
pancreatitis; cirrhosis; glomerular endotheiosis; bacterial infection;
polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
atherosclerosis; cell signal processing-related disorder;
metabolic pathway regulation disorder; cytostatic; neuroprotective;
antiinflammatory; immunosuppressive; analgesic; antihypertensive;
dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;
differentiation; proliferation; motility; haematopoiesis; wound healing;
angiogenesis; forensic biology; transgenic animal; drug screening;
gene therapy; NOV14b; prostatic precursor-like.

XX Homo sapiens.

XX WO200253742-A2.

XX 11-JUL-2002.

XX 07-JAN-2002; 2002MO-US0000375.

XX 05-JAN-2001; 2001US-0260018P.

XX 08-JAN-2001; 2001US-0260360P.

XX 28-FEB-2001; 2001US-0272411P.

XX 02-MAR-2001; 2001US-0272817P.

XX 05-JUL-2001; 2001US-0303231P.

XX 12-JUL-2001; 2001US-0305060P.

XX 10-SEP-2001; 2001US-0318405P.

XX 12-SEP-2001; 2001US-0318700P.

XX 04-JAN-2002; 2002US-00037417.

XX (CURA-) CURAGEN CORP.

XX Kekuda R, Alschbrook JP, Tchernev VT, Liu X, Spytek KA;

PI Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CAM, Li L;

PI	Gorman L, Edinger S, Sciore P, Ellerman K, Malyankar U;	301	TOPQKTSQDCLHOTAFDLSARILLRPLSHISVGVSTGKSLVLPWLSPLHGLWGPF	357
PI	Rochenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;	301	TOPQKTSQDCLHOTAFDLSARILLRPLSHISVGVSTGKSLVLPWLSPLHGLWGPF	357
PI	Padigar M, Taupier R, Miller CE, Eisen A;			
XX				
DR	WPI; 2002-583619/62.			
DR	N-PSDB; ABQ93902.			
XX				
PT	Novel polypeptides and nucleic acids homologous to transmembrane			
PT	receptor, thymosin, neuromodulin-like family of proteins for diagnosing,			
PT	treating cancer, atherosclerosis, neurological, skin and autoimmune			
PT	disorders.			
XX				
PS	Claim 1c; Page 143; 323pp; English.			
XX				
CC	The invention relates to 24 novel human proteins designated NOV1-NOV14			
CC	(ABB09501-ABB09524), collectively referred to as NOVX proteins, and			
CC	nucleic acids encoding them (ABQ93879-ABQ93902). NOVX proteins and			
CC	nucleotides are useful in the treatment, diagnosis or prevention of NOVX-			
CC	associated disorders or in the manufacture of a medicament for treating			
CC	such disorders, with specific applications described for each of the 24			
CC	NOVX proteins, based on their homology to known proteins. Various			
CC	disorders are associated with NOVX proteins including neurological			
CC	disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),			
CC	pain, behavioural disorders, addiction, tuberous sclerosis, cancers			
CC	(e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders			
CC	(e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,			
CC	various forms of arthritis, diabetes, thyroiditis, cardiovascular disease			
CC	(e.g., hypertension), reproductive disorders, endometriosis,			
CC	incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,			
CC	cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine			
CC	disorders, obesity, bacterial infections and particularly cardiomyopathy,			
CC	atherosclerosis, cell signal processing-related disorders and disorders			
CC	of metabolic pathway regulation. NOVX nucleic acids and polypeptides may			
CC	be used to identify cellular receptors or downstream effectors which			
CC	binds to a NOVX protein, and are also useful as targets for the			
CC	identification of small molecules that modulate or inhibit processes such			
CC	as neurogenesis, cell differentiation, cell motility, cellular			
CC	proliferation, haematopoiesis, wound healing and angiogenesis. NOVX			
CC	nucleic acid sequences can be used to identify a cell or tissue type and			
CC	are useful as a source of primers or probes for forensic biology and for			
CC	identifying and cloning NOVX homologues in other cell types. Cells			
CC	comprising NOVX nucleic acids are useful for producing non-human			
CC	transgenic animals which are useful for studying the function and			
CC	activity of NOVX proteins and for identifying and evaluating modulators			
CC	of NOVX activity. The present sequence represents the prostatin precursor			
CC	-like protein NOV14b			
XX				
SQ	Sequence 357 AA;			
	Query Match	100.0%;	Score 1953;	DB 5; Length 357;
	Best Local Similarity	100.0%;	Pred. No. 5.2e-146;	
	Matches 357; Conservative	0;	Mismatches	0; Indels
				0; Gaps
QY	1	MAQKVLGPGQLGAVANSDSLSYLGVSPGARGPPYCGRPEPSARIYVGGNSAQPGTWPW	60	
DB	1	MAQKVLGPGQLGAVANSDSLSYLGVSPGARGPPYCGRPEPSARIYVGGNSAQPGTWPW	60	
QY	61	QVSLHHGGHICGGSLLIAPSWLSAAHCFMTWNTGTEPAEWSVLGVHSQDGLDCAHTR	120	
DB	61	QVSLHHGGHICGGSLLIAPSWLSAAHCFMTWNTGTEPAEWSVLGVHSQDGLDCAHTR	120	
QY	121	AVAAIVVPANYSQVELGADLALLRLASPASLGPVWPVCLPRASHRFVHGTCATWATGMD	180	
DB	121	AVAAIVVPANYSQVELGADLALLRLASPASLGPVWPVCLPRASHRFVHGTCATWATGMD	180	
QY	181	VOEADPLPFWLQVEVELLIGCATCQCLYSQPGFNLTLQILPGMLCAGYEGRRDTCQ	240	
DB	181	VOEADPLPFWLQVEVELLIGCATCQCLYSQPGFNLTLQILPGMLCAGYEGRRDTCQ	240	
QY	241	GDSGGPLVCEGGRWFQAGITSGFCGGRNPGVFTAVATYEAWIREQVMGSEFGPAPP	300	
DB	241	GDSGGPLVCEGGRWFQAGITSGFCGGRNPGVFTAVATYEAWIREQVMGSEFGPAPP	300	

QY	301	TOPQKTSQDCLHOTAFDLSARILLRPLSHISVGVSTGKSLVLPWLSPLHGLWGPF	357
DB	301	TOPQKTSQDCLHOTAFDLSARILLRPLSHISVGVSTGKSLVLPWLSPLHGLWGPF	357
XX			
RESULT 2			
ABB09523			
ID	ABB09523	standard; protein; 344 AA.	
XX			
AC	ABB09523;		
XX			
DT	01-NOV-2002	(first entry)	
XX			
DE	Human prostatin precursor-like NOV14a protein, SEQ ID NO:44.		
XX			
KW	Human; NOVX; neurological disorder; Alzheimer's disease;		
KW	Huntington's disease; Parkinson's disease; pain; behavioural disorder;		
KW	addiction; tuberous sclerosis; cancer; immune disorder; allergy;		
KW	autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;		
KW	thyroiditis; cardiovascular disease; hypertension; reproductive disorder;		
KW	endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;		
KW	pancreatitis; cirrhosis; glomerular endotheliosis; bacterial infection;		
KW	polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;		
KW	atherosclerosis; cell signal processing-related disorder;		
KW	metabolic pathway regulation disorder; cytostatic; neuroprotective;		
KW	antiinflammatory; immunosuppressive; analgesic; antithrombotic;		
KW	dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;		
KW	differentiation; proliferation; motility; haematopoiesis; wound healing;		
KW	angiogenesis; forensic biology; transgenic animal; drug screening;		
KW	gene therapy; NOV14a; prostatin precursor-like; chromosome 16.		
OS	Homo sapiens.		
XX			
XX	W0200253742-A2.		
XX			
PD	11-JUL-2002.		
XX			
XX	07-JAN-2002; 2002WO-US000375.		
XX			
XX	05-JAN-2001; 2001US-0260018P.		
PR	08-JAN-2001; 2001US-0260360P.		
PR	28-FEB-2001; 2001US-0272411P.		
PR	02-MAR-2001; 2001US-0272817P.		
PR	05-JUL-2001; 2001US-0302311P.		
PR	12-JUL-2001; 2001US-0305060P.		
PR	10-SEP-2001; 2001US-0318405P.		
PR	12-SEP-2001; 2001US-0318700P.		
PR	04-JAN-2002; 2002US-00037417.		
XX			
FA	(CURA-) CURAGEN CORP.		
XX			
PI	Kekuda R, Alsobrook JP, Tchernev VT, Liu X, Spytek KA;		
PI	Patturajan M, Grose WM, Lepley DM, Burgess CE, Vernet CAM, Li L;		
PI	Gorman L, Edinger S, Sciore P, Ellerman K, Malyankar U;		
PI	Rochenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;		
PI	Padigar M, Taupier R, Miller CE, Eisen A;		
XX			
DR	WPI; 2002-583619/62.		
DR	N-PSDB; ABQ93901.		
XX			
PT	Novel polypeptides and nucleic acids homologous to transmembrane		
PT	receptor, thymosin, neuromodulin-like family of proteins for diagnosing,		
PT	treating cancer, atherosclerosis, neurological, skin and autoimmune		
PT	disorders.		
XX			
PS	Claim 1c; Page 142; 323pp; English.		
XX			
CC	The invention relates to 24 novel human proteins designated NOV1-NOV14		
CC	(ABB09501-ABB09524), collectively referred to as NOVX proteins, and		
CC	nucleic acids encoding them (ABQ93879-ABQ93902). NOVX proteins and		
CC	nucleotides are useful in the treatment, diagnosis or prevention of NOVX-		
CC	associated disorders or in the manufacture of a medicament for treating		
CC	such disorders, with specific applications described for each of the 24		

CC NOVX proteins, based on their homology to known proteins. Various
CC disorders are associated with NOVX proteins including neurological
CC disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
CC pain, behavioural disorders, addiction, tuberosclerosis, cancers
CC (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
CC (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,
CC various forms of arthritis, diabetes, thyroiditis, cardiovascular disease
CC (e.g., hypertension), reproductive disorders, endometriosis,
CC cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
CC disorders, obesity, bacterial infections and particularly cardiomyopathy,
CC atherosclerosis, cell signal processing-related disorders and disorders
CC of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
CC be used to identify cellular receptors or downstream effectors which
CC binds to a NOVX protein, and are also useful as targets for the
CC identification of small molecules that modulate or inhibit processes such
CC as neurogenesis, cell differentiation, cell motility, cellular
CC proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
CC nucleic acid sequences can be used to identify a cell or tissue type and
CC are useful as a source of primers or probes for forensic biology and for
CC identifying and cloning NOVX homologues in other cell types. Cells
CC comprising NOVX nucleic acids are useful for producing non-human
CC transgenic animals which are useful for studying the function and
CC activity of NOVX proteins and for identifying and evaluating modulators
CC of NOVX activity. The present sequence represents the prostatic precursor
CC -like protein NOV14a. The gene encoding NOV14a is located on chromosome
CC 16
XX
SQ Sequence 344 AA;

Query Match 96.0%; Score 1874; DB 5; Length 344;
Best Local Similarity 100.0%; Pred. No. 8.6e-140;
Matches 344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAQKVLGPGQLGAVANSYSYLGVPSPGAPGPPYCGRPSPARIVGGSNAQPGTWP 60
Db 1 MAQKVLGPGQLGAVANSYSYLGVPSPGAPGPPYCGRPSPARIVGGSNAQPGTWP 60
QY 61 QVSLHGGGHHICGSLIAPSWLSAAHCFWNTGTLPEAAEWSVLLGVHSQDGLDGAHTR 120
Db 61 QVSLHGGGHHICGSLIAPSWLSAAHCFWNTGTLPEAAEWSVLLGVHSQDGLDGAHTR 120
QY 121 AVAAIIVVPANYSQVELGADLALLRLASPSLGSAPVAVPVCPLPRASHRFVHGTCWATGWGD 180
Db 121 AVAAIIVVPANYSQVELGADLALLRLASPSLGSAPVAVPVCPLPRASHRFVHGTCWATGWGD 180
QY 181 VQEAADPLPLPWLQVLELLGEATCQCLYSQPGPENLTQLPLGMLCAGYEGREDTCQ 240
Db 181 VQEAADPLPLPWLQVLELLGEATCQCLYSQPGPENLTQLPLGMLCAGYEGREDTCQ 240
QY 241 GDSGGPLVCEEGRWPFQAGITSFQCGRRNRPVFTAVATYEAWIREQVMGSEPGAPFP 300
Db 241 GDSGGPLVCEEGRWPFQAGITSFQCGRRNRPVFTAVATYEAWIREQVMGSEPGAPFP 300
QY 301 TOPQKTSQCLHQTAFDLSARILLRPLSHISVGVSTGTSKSLVLP 344
Db 301 TOPQKTSQCLHQTAFDLSARILLRPLSHISVGVSTGTSKSLVLP 344

RESULT 3
ABB98135
ID ABB98135 standard; protein; 818 AA.
XX
AC ABB98135;
XX
DT 17-OCT-2002 (first entry)
XX
DE Human PMMM Incyte ID 2751509CD1.
XX
KW Human; PMMM; protein modification and maintenance molecule;
KW anticonvulsant; neuroprotective; nootropic; cytostatic; antipsoriatic;
KW antiasthmatic; dermatological; antidiabetic; antiparkinsonian;
KW antiataxic; antiinflammatory; antiulcer; antiangular; cardiac;

KW hepatotropic; osteopathic; antiemetic; antipyretic; virucide;
KW antibacterial; fungicide; gastrointestinal; antidiarrheic; laxative;
KW haemostatic; anti-HIV; thrombolytic; anticoagulant; gynaecological;
KW cardiovascular; antiarteriosclerotic; hypotensive; vasotropic;
KW antitumor; antihemetic; immunosuppressive; antiallergic; antithyroid;
KW nephrotropic; antigout; thyromimetic; antiarthritic; uropathic;
KW ophthalmological; antiparasitic; tranquiliser; vulnerary; keratolytic;
KW auditory; antiseborrheic; antidepressant; neuroleptic; antiinfertility;
KW antelmintic; protozoacide; Crohn's disease; hypertension; autoimmune;
KW inflammatory; anaemia; cell proliferative; developmental; epithelial;
KW scabies; neurological; Alzheimer's disease; reproductive;
KW ectopic pregnancy; gene therapy; vaccine; disorder; prostasin.
OS Homo sapiens.
PN WO200246383-A2.
XX
PD 13-JUN-2002.
XX
OS 05-DEC-2001; 2001WO-US046964.
XX
PR 08-DEC-2000; 2000US-0254399P.
PR 21-DEC-2000; 2000US-0257803P.
PR 05-JAN-2001; 2001US-0260110P.
PR 19-JAN-2001; 2001US-0262851P.
PR 25-JAN-2001; 2001US-0264623P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Azimzai Y, Kallick DA, Baughn MR, Griffin JA, Swarnakar A;
PI Lal PG, Wallia NK, Hafalia AJA, Gandhi AR, Au-Young J, Elliott VS;
PI Ramkumar J, Thangavelu K, Lu Y, Warren BA, Lu DAM, Lee EA;
PI Tribouley CM, Arvizu C, Delegeane AM, Yao MG, Khan FA;
PI Sanjanwala MM;
XX
DR WPI; 2002-519664/55.
XX
DR N-PSDB; ABQ75956.
XX
PT New isolated Protein Modification and Maintenance polypeptides, useful
PT for diagnosis, and treatment of e.g. gastrointestinal disorders.
XX
PS Claim 1 (a); Page 174-176; 200pp; English.
XX
CC The invention relates to an isolated Protein Modification and Maintenance
CC (PMM) polypeptide. Polypeptides of the invention may be used in the
CC diagnosis, treatment and prevention of disorders associated with
CC decreased expression or activity of PMM. These include gastrointestinal
CC disorders (e.g. Crohn's disease), cardiovascular disorders (e.g. hypertension), autoimmune/inflammatory disorders (e.g. anemias), cell
CC proliferative disorders, developmental disorders, epithelial disorders
CC (e.g. scabies), neurological disorders (e.g. Alzheimer's disease)
CC reproductive disorders (e.g. ectopic pregnancy), and in gene therapy or a
CC vaccine for such diseases. They may also be used in the assessment of the
CC effects of exogenous compound on the expression of nucleic acid and amino
CC acid sequences of protein modification and maintenance molecules. The
CC current sequence represents a human PMM of the invention, which has been
CC found to have homology with rat prostasin
XX
SQ Sequence 818 AA;

Query Match 77.1%; Score 1505; DB 5; Length 818;
Best Local Similarity 100.0%; Pred. No. 3.2e-110;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 38 CGRPEPSARIIVGGSNAQPGTWPQVSLHHGGHHICGSLIAPSWLSAAHCFWNTGTLPE 97
Db 38 CGRPEPSARIIVGGSNAQPGTWPQVSLHHGGHHICGSLIAPSWLSAAHCFWNTGTLPE 97
QY 98 AAEEWSVLLGVHSQDGLDGAHTRVAAIIVVPANYSQVELGADLALLRLASPSLGSAPVAVP 157
Db 98 AAEEWSVLLGVHSQDGLDGAHTRVAAIIVVPANYSQVELGADLALLRLASPSLGSAPVAVP 157
QY 158 VCLPRASHRFVHGTCWATGWGDVQEAADPLPLPWLQVLELLGEATCQCLYSQPGPEN 217


```

CC format directly from EPO
XX Sequence 766 AA;
SQ

Query Match      61.5%; Score 1202; DB 4; Length 766;
Best Local Similarity 100.0%; Pred. No. 2.5e-86;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 MTNGTLEPAAEWSVLLGVHSQDGLDGAHTRAVAAIVVPANYSQVELGADLALLRLASPA 149
Db 1 MTNGTLEPAAEWSVLLGVHSQDGLDGAHTRAVAAIVVPANYSQVELGADLALLRLASPA 60

QY 150 SLGPAPVPCLPRAHFRVHGTACWATGWDVQVADPLPLPWVQLQVELRLGECATCQCL 209
Db 61 SLGPAPVPCLPRAHFRVHGTACWATGWDVQVADPLPLPWVQLQVELRLGECATCQCL 120

QY 210 YSQPGPNLTQLLPGMLCAGYPEGRRDTCQDGGGLVCEEGRWFGAGITSFGGCGR 269
Db 121 YSQPGPNLTQLLPGMLCAGYPEGRRDTCQDGGGLVCEEGRWFGAGITSFGGCGR 180

QY 270 RNRPGVFTAVATYEATREQVWMSSEPGAPPTQPKTQSD 309
Db 181 RNRPGVFTAVATYEATREQVWMSSEPGAPPTQPKTQSD 220

RESULT 6
AAU75082
ID AAU75082 standard; protein; 178 AA.
XX
AC AAU75082;
XX
08-MAY-2002 (first entry)
XX
Human prostatic-like serine protease protein.
XX
Human; prostatic-like serine protease; cytostatic; antiatherosclerotic;
KW viricide; osteopathic; antiinflammatory; vasotropic; neuroprotective;
KW trypsin-like; metastasis; autoimmune lesion; atherosclerosis;
KW renal failure; malignant cell; tumour angiogenesis; enzyme; inflammation;
KW chronic obstructive pulmonary disease; COPD; restenosis; atherosclerosis;
KW neurodegenerative disease; prion protein; infection; amyloid plaque;
KW Genetmann-Straussler Syndrome; viral infection; Scrapie;
KW Creutzfeldt-Jakob disease; metastatic cancer; lipid accumulation;
XX osteoporosis; Paget's disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 1..14 /note= "Kringles domain"
FT 1..13 /note= "Serine protease trypsin family domain"
FT 8..12 /note= "Trypsin_His region"
FT 72..108 /note= "Type I fibronectin domain"
FT 83..103 /note= "Kringles domain"
FT 154..177 /note= "Apple protein domain"
FT 161..174 /note= "Type I fibronectin domain"
FT 162..177 /note= "Serine protease trypsin family domain"
FT 162..173 /note= "Trypsin_ser region"
XX
WO200198467-A2.
XX
27-DEC-2001.
XX
22-JUN-2001; 2001WO-BF007117.
XX

```

```

PR 23-JUN-2000; 2000US-0213588P.
PR 20-MAR-2001; 2001US-0276909P.
XX (FARB ) BAYER AG.
XX
XX Xiao Y, Morozov V;
XX
XX MPI; 2002-114576/15.
XX N-PSDB; ABK13565.
XX
PT Novel human prostatic-like serine protease polypeptide and polynucleotide
PT which can be regulated for treating metastasis of malignant cells,
PT inflammation, atherosclerosis, neurodegenerative disease and infections.
XX
XX Claim 25; Fig 2; 111pp; English.
XX
XX This invention comprises the cDNA and protein sequences of an isolated
XX prostatic-like serine protease and reagents and methods for regulating
XX the human prostatic-like enzyme activity. Prostatic is a trypsin-like
XX serine protease purified from human seminal fluid. An antibody specific
XX for prostatic-like serine protease is useful for immunodetection and
XX diagnosis of micro-metastases, autoimmune lesions and renal failure in
XX biopsy specimens, plasma samples and body fluids. The antibody may be
XX used to modulate enzyme activity in a disease, such as metastasis of
XX malignant cells, tumour angiogenesis, inflammation, chronic obstructive
XX pulmonary disease (COPD), atherosclerosis, neurodegenerative disease
XX (e.g. prion protein amyloid plaques of Genetmann-Straussler Syndrome,
XX Creutzfeldt-Jakob disease, Scrapie) or infection, particularly viral
XX infection. The human prostatic-like serine protease gene provides a
XX therapeutic target of decreasing human prostatic-like serine protease
XX activity, in particular for treating or preventing metastatic cancer. The
XX agonists and antagonists of the nucleotide sequence may be used to treat
XX osteoporosis, Paget's disease and degradation of bone implants
XX particularly dental implants. Altered levels of human prostatic-like
XX serine protease activity inhibits both smooth muscle cell proliferation
XX and lipid accumulation and inhibit the progression of restenosis and
XX atherosclerosis. The nucleic acid sequence is also useful in diagnostic
XX assays for detecting diseases and abnormalities or susceptibility to
XX diseases related to the presence of mutations in nucleic acid sequences
XX which encode the enzyme. The present sequence represents the human
XX prostatic-like serine protease sequence of the invention
XX
XX Sequence 178 AA;
SQ
Query Match      49.7%; Score 970; DB 5; Length 178;
Best Local Similarity 100.0%; Pred. No. 9.9e-69;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 LIAPSWVLSAHCFTMTNGTLEPAAEWSVLLGVHSQDGLDGAHTRAVAAIVVPANYSQVE 135
Db 1 LIAPSWVLSAHCFTMTNGTLEPAAEWSVLLGVHSQDGLDGAHTRAVAAIVVPANYSQVE 60

QY 136 LGADLALLRLASPAHFRVHGTACWATGWDVQVADPLPLPWVQLQVELRLGECATCQCL 195
Db 61 LGADLALLRLASPAHFRVHGTACWATGWDVQVADPLPLPWVQLQVELRLGECATCQCL 120

QY 196 VELRLGECATCQCLYSQPGPNLTQLLPGMLCAGYPEGRRDTCQDGGGLVCEEGG 253
Db 121 VELRLGECATCQCLYSQPGPNLTQLLPGMLCAGYPEGRRDTCQDGGGLVCEEGG 178

RESULT 7
ABG04853
ID ABG04853 standard; protein; 198 AA.
XX
XX ABG04853;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #4844.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX

```

KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS69040.

XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

PS Claim 20; SEQ ID NO 35212; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 198 AA;

Query Match 41.2%; Score 804.5; DB 4; Length 198;
Best Local Similarity 96.2%; Pred. No. 1.3e-55;
Matches 150; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 93 GTLEPAEWSVLGVHSGQDPLDGAHTRAVAAIVVPANTYSQVELGADIALRLASPASIG 152

Db 1 GTLEPAEWSVLGVHSGQDPLDGAHTRAVAAIVVPANTYSQVELGADIALRLASPASIG 60

QY 153 PAVWPVCLPRASHRFVHGTAACWATGWDVQVADPLPLPWLOVEVRLLEATCCCLYSQ 212

Db 61 PAVWPVCLPRASHRFVHGTAACWATGWDVQVADPLPLPWLOVEVRLLEATCCCLYSQ 120

QY 213 PGPFNLTLQLFGLMGLCAGYEGEGRDTCQ-GDSGGPL 247

Db 121 PGPFNLTLQLFGLMGLCAGYEGEGRDTCQSPERGPM 156

RESULT 8

ABG96402

ID ABG96402 standard; protein; 343 AA.

XX AC ABG96402;

XX

DT 11-DEC-2002 (first entry)

XX Human ovarian cancer marker OV80.

XX Human; ovarian cancer; marker; cancer; familial history; brain disorder; central nervous system disorder; bacterial meningitis; viral meningitis; Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus; brain herniation; inflammation; encephalitis; testicular disorder; nontuberculous granulomatous orchitis; connective tissue disorder; heart disorder; ischaemic heart disease; atherosclerosis; neoplasm; histological type; carcinogenic; ovarian cancer marker.

OS Homo sapiens.

XX WO200271928-A2.

PN 19-SEP-2002.

XX 14-MAR-2002; 2002WO-US007826.

XX 14-MAR-2001; 2001US-0276025P.

PR 14-MAR-2001; 2001US-0276036P.

PR 10-AUG-2001; 2001US-0311732P.

PR 19-SEP-2001; 2001US-0323580P.

PR 26-SEP-2001; 2001US-0324967P.

PR 26-SEP-2001; 2001US-0325102P.

PR 26-SEP-2001; 2001US-0325149P.

XX (MILL-) MILLENNIUM PHARM INC.

PI Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;

PI Meyers RE, Morrissey MP, Olandt PJ, Sen A, Vieby PO, Mills GB;

PI Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;

XX WPI; 2002-723277/78.

DR N-PSDB; ABS76501.

XX Assessing whether a patient is afflicted with ovarian cancer, useful in assessing the stage or progression of the disease, comprises comparing the expression level of a cancer marker in a sample from a patient and from a non cancer patient.

PS Disclosure; Page 385; 481pp; English.

CC The present invention relates to a new method for assessing whether a patient is afflicted with ovarian cancer. The method involves comparing the expression level of a marker in a patient sample and the normal level of expression of the marker in a control non-ovarian cancer sample, where the marker is selected from 363 cancer markers described in the specification. The method of the invention is useful in diagnosing or characterising cancer, in detecting the presence of cancer as early as possible, and the recurrence of ovarian cancer. The method may also be of particular use with patients having an enhanced risk of developing ovarian cancer (e.g. patients having a familial history of ovarian cancer). The cancer markers may be used in the management and treatment of e.g. brain and central nervous system disorders (e.g. bacterial and viral meningitis, Alzheimer's disease or Parkinson's disease), brain disorders (e.g. cerebral oedema, hydrocephalus or brain herniations), inflammations (e.g. bacterial or viral meningitis or encephalitis), testicular disorders (e.g. nontuberculous granulomatous orchitis), connective tissue disorders, or heart disorders (e.g. ischaemic heart disease or atherosclerosis). The compositions and methods may also be used in assessing the histological type of neoplasm associated with ovarian cancer, monitoring the progression of ovarian cancer, determining whether ovarian cancer has metastasized or is likely to metastasize, selecting a composition for inhibiting ovarian cancer, assessing the ovarian carcinogenic potential of a compound, or inhibiting ovarian cancer or at risk of developing ovarian cancer. The present amino acid sequence represents one of the ovarian cancer markers described in the invention

XX Sequence 343 AA;

SQ

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 343 AA;
 Query Match 35.3%; Score 690; DB 7; Length 343;
 Best Local Similarity 43.6%; Pred. No. 2.7e-46;
 Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;
 QY 1 MAQKGVLPQGLGAVANSYSYLY-GLVPSGP-ARGPPYCGRPEPSARIVGGSNAQP 55
 DB 1 MAQKGVLPQGLGAVA---ILLYLGLLSRGTGAEGAEAP--CG-VAPOARITGSSAVA 53
 QY 56 GTWPMQVSLHGGHICGSLTAPSVLSAAHCFMTNGTLEPAEWSVLLGVHSDGDPD 115
 DB 54 GQWPQVQSVITGVHVCGLSLVSEQWLSAAHCFSEHKE--AYEVKIGAHQDSYSE 110
 QY 116 GAHTRAVAAIVVPANYSQVELGADLALLRLASPAISGPAVVPVCLPRASHRFVHGTACWA 175
 DB 111 DAKVSTLKDIIHPHPSVLOEGSQDIALQLSRPITFSYRIRPCLPAAANSPFNGLHCTV 170
 QY 176 TCGDVGQADPLPLPWLQVEVLRLGATCOCLYSQPGFNLTLQILPGLMCAQYVPEGR 235
 DB 171 TCGHVPASVSLTPKPLQQLVPLISRETCLNIDAKPEEPHFVQEDMVCAGYVEGG 230
 QY 236 RDTCCGDSGGPLVCEBGGFWOAGITSGFGRNRPGVFTAVATYEAWIREQVMGSEPR 295
 DB 231 KDACQDSGGPLSCPVEGLWYLTGIVSWGDCAGARNRPGVYTLASSYASWIOSKV--TEL 288
 QY 296 GPAPFTQPKTQSD---CLHQTAFDLS-ARILLRPLSHISVGVSTGKSLVLPWLSPH 349
 DB 289 QPRVVPQTQESQPSDNLCSHLAFSSAPAQGLRPLFLPLGLALG---LLSPWLSEH 343

RESULT 14
 ABB98415
 ID ABB98415 standard; protein; 307 AA.
 AC ABB98415;
 XX 21-OCT-2002 (first entry)
 DT Human NOV14a, prostaticin-like protein.
 DE Human NOV14a, prostaticin-like protein.
 XX Human; NOV14a; cytostatic; Cardiant; Antiinflammatory; Immunosuppressive;
 KW Antiallergic; Haemostatic; Anti-HIV; Antidiabetic; Anorectic;
 KW Antiasthmatic; Nephrotropic; Hepatotropic; Neuroprotective; Nootropic;
 KW Antibacterial; Virucide; Antiparasitic; Relaxant; Anticonvulsant;
 KW Gene Therapy; NOV; cancer; heart disease; inflammation;
 KW autoimmune disorder; allergy; blood disorder; AIDS; diabetes; obesity;
 KW asthma; IGA nephropathy; cirrhosis; arthritis; Alzheimer's disease;
 KW infection; stroke; muscular dystrophy; epilepsy; wasting disorder;
 KW prostaticin-like protein.
 XX Homo sapiens.
 OS
 PN WO200255704-A2.
 XX 18-JUL-2002.
 XX 09-JAN-2002; 2002WO-05000554.
 XX 09-JAN-2001; 2001US-0260417P.
 PR 10-JAN-2001; 2001US-0260831P.
 PR 28-FEB-2001; 2001US-0272338P.
 PR 09-MAR-2001; 2001US-0274876P.
 PR 18-APR-2001; 2001US-0284704P.
 XX (CURA-) CURAGEN CORP.
 PA Padigar M, Li L, Zerhusen BD, Casman SJ, Shenoy S, Spytek KA;
 PI Zhong M, Gangolli EA, Burgess CE, Fatturajan M, Vernet CAM;

PI Taylor S, Tchernev VT, Miller CE, Guo X, Boldog FL, Grosse WM;
 PI Alsbrook JP, Gerlach V, Edinger S, Rothenberg ME, Ellerman K;
 PI MacDougall J, Malyankar U, Millet I, Peyman J, Smithson G;
 PI Gunther E, Stone DU;
 XX WPI; 2002-590674/63.
 DR N-PSDB; ABB85392.
 XX NOVX polypeptides and encoding polynucleotides, useful for preventing or
 PT treating NOVX-associated disorders e.g. cancer, inflammation, or
 PT Alzheimer's disease, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.
 XX Claim 1; Page 98; 358pp; English.

CC The present sequence is the protein sequence for a NOV protein. The NOV
 CC proteins and coding sequences are useful for treating or preventing NOV-
 CC associated disorders or in the manufacture of a medicament for treating
 CC the disorders, such as cancer, heart disease, inflammation, autoimmune
 CC disorders, allergies, blood disorders, AIDS, diabetes, obesity, asthma,
 CC IGA nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections
 CC (e.g. bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy,
 CC and other wasting disorders associated with chronic diseases. NOV14a is a
 CC prostaticin-like protein, and the NOV14a coding sequence localises to
 CC Chromosome 16p11.2

XX Sequence 307 AA;
 Query Match 33.0%; Score 645; DB 5; Length 307;
 Best Local Similarity 41.3%; Pred. No. 8.5e-43;
 Matches 148; Conservative 45; Mismatches 105; Indels 60; Gaps 11;

QY 1 MAQKGVLPQGLGAVANSYSYLY-GLVPSGP-ARGPPYCGRPEPSARIVGGSNAQP 55
 DB 1 MAQKGVLPQGLGAVA---ILLYLGLLSRGTGAEGAEAP--CG-VAPOARITGSSAVA 53
 QY 56 GTWPMQVSLHGGHICGSLTAPSVLSAAHCFMTNGTLEPAEWSVLLGVHSDGDPD 115
 DB 54 GQWPQVQSVITGVHVCGLSLVSEQWLSAAHCF-----87
 QY 116 GAHTRAVAAIVVPANYSQVELGADLALLRLASPAISGPAVVPVCLPRASHRFVHGTACWA 175
 DB 88 -----PSEHHKGSQ-DIALQLSRPTSYSYRIRPCLPAAANSPFNGLHCTV 134
 QY 176 TCGDVGQADPLPLPWLQVEVLRLGATCOCLYSQPGFNLTLQILPGLMCAQYVPEGR 235
 DB 135 TCGHVPASVSLTPKPLQQLVPLISRETCLNIDAKPEEPHFVQEDMVCAGYVEGG 194
 QY 236 RDTCCGDSGGPLVCEBGGFWOAGITSGFGRNRPGVFTAVATYEAWIREQVMGSEPR 295
 DB 195 KDACQDSGGPLSCPVEGLWYLTGIVSWGDCAGARNRPGVYTLASSYASWIOSKV--TEL 252
 QY 296 GPAPFTQPKTQSD---CLHQTAFDLS-ARILLRPLSHISVGVSTGKSLVLPWLSPH 349
 DB 253 QPRVVPQTQESQPSDNLCSHLAFSSAPAQGLRPLFLPLGLALG---LUSPWLSEH 307

RESULT 15
 ABB98415
 ID ABB98415 standard; protein; 339 AA.
 XX AC ABB98415;
 XX 24-JAN-2003 (first entry)
 DT Mouse channel activating protease 1 (CAP1) protein.
 DE Mouse channel activating protease 1 (CAP1) protein.
 XX Mouse; enzyme; transgenic; channel activating protease 1; CAP1;
 KW serine protease; seizure; epilepsy; therapeutic; agonist; anticonvulsant.
 XX Mus musculus.
 OS
 PN
 XX Key Location/Qualifiers
 FH

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 15:37:18 ; Search time 45 Seconds
(without alignments)
763.119 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKGVLPGLGAVANSDS.....TKSLVPLSPHSLILGLMGF 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	690	35.3	343	1 A57014	proctasin (EC 3.4.21.1)
2	530	27.1	855	2 JCT731	membrane-bound arg
3	514.5	26.3	275	2 A32410	tryptase (EC 3.4.21.1)
4	499	25.6	276	2 A38654	mast cell protease
5	496	25.4	1019	1 A56318	enteropeptidase (E
6	491.5	25.2	274	2 JCA171	tryptase (EC 3.4.21.1)
7	490.5	25.1	417	1 S00845	hepsin (EC 3.4.21.1)
8	485.5	24.9	275	2 A35863	tryptase (EC 3.4.21.1)
9	484.5	24.8	275	2 B35863	tryptase (EC 3.4.21.1)
10	482.5	24.7	638	1 KOHUP	plasma kallikrein
11	481.5	24.7	274	2 A45754	tryptase (EC 3.4.21.1)
12	480	24.6	1034	1 A53663	enteropeptidase (E
13	478.5	24.5	275	2 C35863	tryptase (EC 3.4.21.1)
14	478.5	24.5	416	1 S33777	hepsin (EC 3.4.21.1)
15	475	24.3	638	1 KOMSPL	plasma kallikrein
16	474.5	24.3	270	2 S56160	mast cell tryptase
17	473.5	24.2	418	2 A37344	acrosin (EC 3.4.21.1)
18	473	24.2	421	2 S29599	acrosin (EC 3.4.21.1)
19	473	24.2	638	1 KQRTPL	plasma kallikrein
20	469.5	24.0	1035	1 A43090	enteropeptidase (E
21	469	24.0	436	2 JX0172	acrosin (EC 3.4.21.1)
22	467.5	23.9	431	2 S47538	acrosin (EC 3.4.21.1)
23	465.5	23.8	273	2 A47246	tryptase (EC 3.4.21.1)
24	464.5	23.8	269	2 B32410	mastocytoma protei
25	464.5	23.8	612	1 PLBO	plasmin (EC 3.4.21.1)
26	462	23.7	415	1 A34170	acrosin (EC 3.4.21.1)
27	460.5	23.6	437	2 S18407	acrosin (EC 3.4.21.1)
28	454	23.2	455	2 A61545	plasmin (EC 3.4.21.1)
29	452	23.1	460	2 B61545	plasmin (EC 3.4.21.1)

low-density lipopr
acrosin (EC 3.4.21.1)
plasmin (EC 3.4.21.1)
coagulation factor
plasmin (EC 3.4.21.1)
plasmin (EC 3.4.21.1)
plasmin (EC 3.4.21.1)
coagulation factor
plasmin (EC 3.4.21.1)
serine proteinase
polyprotein - Afri
pancreatic elastas
pancreatic elastas
trypsin (EC 3.4.21.1)
testicular serine
acrosin (EC 3.4.21.1)

ALIGNMENTS

RESULT 1

A57014 proctasin (EC 3.4.21.1) precursor - human
C:Species: Homo sapiens (man)
C:Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 21-Apr-2003
C:Accession: A57014; A54866
R:Yu, J.X.; Chao, L.; Chao, J.
J. Biol. Chem. 270, 13483-13489, 1995
A:Title: Molecular cloning, tissue-specific expression, and cellular localization of huma
A:Reference number: A57014; MUID:95286644; PMID:7768952
A:Accession: A57014
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-343 <RES>
A:Cross-references: GB:L41351; NID:9862304; PIDN:AAC41759.1; PID:9862305
A:Experimental source: prostate
A:Note: parts of this sequence were determined by protein sequencing
R:Yu, J.X.; Chao, L.; Chao, J.
J. Biol. Chem. 269, 18843-18848, 1994
A:Title: Proctasin is a novel human serine proteinase from seminal fluid. Purification, t
A:Reference number: A54866; MUID:94308140; PMID:8034638
A:Accession: A54866
A:Molecule type: protein
A:Residues: 45-64 <YUA>
C:Genetics:
A:Gene: GDB:PRSS8
A:Cross-references: GDB:676446; OMIM:600823
A:Map position: 16p11.2-16p11.2
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-44/Domain: proctasin light chain #status predicted <CHL>
F:33-44/Domain: proctasin heavy chain #status predicted <CHH>
F:45-281/Domain: trypsin homology <TRY>
F:323-341/Domain: transmembrane #status predicted <TMML>
F:37-154/Domain: transmembrane #status predicted <SIG>
F:85.134.236/Active site: His, Asp, Ser #status predicted
F:159/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 35.3%; Score 690; DB 1; Length 343;
Best Local Similarity 43.6%; Pred. No. 1.1e-46;
Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;

QY 1 MAQKGVLPGLGAVANSDSYSLY-GLVPSP-----ARGPYCGRPSPSARIVGSSNAQP 55
DB 1 MAQKGVLPGLGAVANSDSYSLY-GLVPSP-----ARGPYCGRPSPSARIVGSSNAQP 53
QY 56 GTWFWQVSLHGGHICGSLIAPSWLSAAHCFMTNCTLEPAAEWSVLLGVHSDGGLD 115
DB 54 GQFWQVSLHGGHICGSLIAPSWLSAAHCFMTNCTLEPAAEWSVLLGVHSDGGLD 110

```

QY 116 GAHTRAVAAIVVPANYSQVELGADLALLRLASPASLGPAVMPVCLPRASHRFVHGTCWA 175
Db 111 DAKVSTLKDIIIPHPSYLQEGSQDIALQLGRPIITFSYIRPICLPANAFPPNGLHCTV 170
QY 176 TQWGVQVQADPLPLPWVQVLESLRLGATCCQLYSQSPGPNLTLQILFGMLCAGYPEGR 235
Db 171 TQWGHVAVSVSLTKPKLQLEVLPLSRETCLNIDAKPEEPHFVQEDMVCAGYVGG 230
QY 236 RDTCCGDSGGLPVCEBEGGRWFQAGITSGFGGRRNRPGVFTAVATYPAWIKREQVMGSEP 295
Db 231 KDACCQDSGGLPSCPVEGLWYLTGIVSGDAGCAENRPGVYTLASSYASWIOSKV--TEL 288
QY 296 GFAFPPTQPKTQSD---CLHQTAFLDS-ARILLRLSHLSVGVSTGKSLVLPWLSPH 349
Db 289 QPRVVPQTQESQDNLCSHLAFSSAPAGQLRLPLFLGLALG---LLSPWLSH 343

RESULT 2
JC7731
membrane-bound arginine-specific serine proteinase precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 03-Feb-2003
C:Accession: JC7731; JC7775
R:Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Athauda,
J. Biochem. 130, 425-430, 2001
A:Title: Characterization of a membrane-bound arginine-specific serine proteinase from rat
A:Reference number: JC7731; MUID:21421307; PMID:11530019
A:Accession: JC7731
A:Molecule type: mRNA
A:Residues: 1-855 <KIS>
A:Cross-references: DDBJ:AB049189
A:Experimental source: Strain Male, 7-week-old
R:Satom, S.; Yanasaki, Y.; Tsuzuki, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.
Biochem. Biophys. Res. Commun. 287, 995-1002, 2001
A:Title: A role for membrane-type serine protease (MT-SPI) in intestinal epithelial turn
A:Reference number: JC7775; PMID:11573963
A:Contents: Small intestine
A:Accession: JC7775
A:Molecule type: mRNA
A:Residues: 1-855 <SAT>
A:Cross-references: DDBJ:AB037898
C:Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease
of specific proteins or peptides on the brushborder membranes. It also participates in
lial migration and/or cell loss.
C:Genetics:
A:Gene: mt-spi
A:Map position: basolateral cell surface
C:Superfamily: membrane-bound arginine-specific serine proteinase
C:Keywords: protein digestion

Query Match 27.1%; Score 530; DB 2; Length 855;
Best Local Similarity 39.9%; Pred. No. 1e-33;
Matches 112; Conservative 40; Mismatches 95; Indels 34; Gaps 8;

QY 33 RGPVYC-GRPEPS-----ARIVGSSNAQPGTWPQVSLHH-GGGHIC 72
Db 582 KGNPECDKDKDCSGSDEKNCDCGLRSFTKQARVVGGINADEGEPWQVSLHALGQGHLC 641
QY 73 GGSIIAPSWLSAAHCFMTNGLTLEPA--AEWSVLGVHSQD-GPLDGHATFAVAIVVPA 129
Db 642 GASLISPDWLSAAHCFQDETIFYSDHTMTWTAFLGLDDQSKRSASGVQEHKRLIITHP 701
QY 130 NYSQVELGADLALLRLASPASLGPAVMPVCLPRASHRFVHGTCWATGWDVQADPLPL 189
Db 702 SFNDFTFDYDIALLEKPAVSTVVRPICLPDNTHTVPAGKALVWTGNGHTEGGTGAL 761
QY 190 PWVQVELRLGELATCCQLYSQSPGPNLTLQILFGMLCAGYPEGRDTCQDSDGGPLV 248
Db 762 --ILQKGIRVINQTTCELLPQ-----QITPRMVCVGLSGVDSCQDSDGGPLSS 811
QY 249 CEEGRWFQAGITSGFGGRRNRPGVFTAVATYPAWIKREQ 289
Db 812 VEKDGRIQAGVSVGEGCAQRNKPQVYTRIPVDRDIKEQ 852

```

```

RESULT 3
A32410
tryptase (EC 3.4.21.59) precursor - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 22-Jun-1999
C:Accession: A32410
R:Vanderslice, P.; Craik, C.S.; Nadel, J.A.; Caughey, G.H.
Biochemistry 28, 4148-4155, 1989
A:Title: Molecular cloning of dog mast cell tryptase and a related protease: structural
A:Reference number: A32410; MUID:89352460; PMID:2504277
A:Accession: A32410
A:Molecule type: mRNA
A:Residues: 1-275 <VAN>
A:Cross-references: GB:M24664; NID:g163982; PIDN:AAA30854.1; PID:g163983; GB:J02862
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-30/Domain: activation peptide #status predicted <ACT>
F:31-275/Product: tryptase #status predicted <MAT>
F:31-267/Domain: trypsin homology <TRY>
F:74,121,224/Active site: His, Asp, Ser #status predicted
Query Match 26.3%; Score 514.5; DB 2; Length 275;
Best Local Similarity 42.2%; Pred. No. 4.8e-33;
Matches 124; Conservative 27; Mismatches 100; Indels 43; Gaps 9;

QY 23 LYGLVPSGARGPPYCGRPEPSARIVGSSNAQPGTWPQVSLHHGG---GHICGSLIAP 79
Db 12 LGSLLVEVSPAP-----GQALQKRVGIVGGREAPCGSKWPQVSLKQYWRHICGSLIHP 66
QY 80 SWVLSAAHCFMTNGLTLEPAAEWSVLL---GVHSQDGLDGAHTRAVAAIVVPANYSQVEL 136
Db 67 QNVLTAAHCVGN--VVCPEELRVQLREQHLYQD-----HLLPVRIVHPNYYTPEN 118
QY 137 GADLALLRLASPASLGPAVMPVCLPRASHRFVHGTCWATGWDVQADPLPLPWVQV 196
Db 119 GADIALLELDEFVNSAHVQVTLPLPQALQTFPTGTCVWTGWDVHSGTLPPLPPFLKQV 178
QY 197 ELRLGELATCCQLYSQSPGPNLTLQILPG-----MLCAGYPEGRDTCQDSDGGPLVC 249
Db 179 KPIVENSMDVQY-----HLGLSGDGVIRVREDMLCAG--NSKSDSCQDSDGGPLVC 230
QY 250 EEGGRWFQAGITSGFGGRRNRPGVFTAVATYPAWIKREQVMGSEPFPPTQ 303
Db 231 RVRGVWLQAGVSVGEGCAQRNPGIYTRVAYVLDWIHQVY-----PKEP 275

RESULT 4
A38654
mast cell proteinase 6 (EC 3.4.21.-) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-Feb-1992 #sequence_revision 17-Feb-1994 #text_change 22-Jun-1999
C:Accession: A38654; B38654; D35646; I59478
R:Reynolds, D.S.; Gurley, D.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 266, 3847-3853, 1991
A:Title: Cloning of the cDNA and gene of mouse mast cell protease-6. Transcription by p
A:Reference number: A38654; MUID:91139682; PMID:1995638
A:Accession: A38654
A:Molecule type: DNA
A:Residues: 1-276 <REY>
A:Cross-references: GB:M57625; NID:g200506; PIDN:AAA39987.1; PID:g200507
A:Note: The authors translated the codon CGC for residue 24 as Ala, and GAA for residue 37
s Gly, GAG for residue 148 as Gly, GAG for residue 168 as Gly, and GAA for 185 as Gly
A:Accession: B38654
A:Molecule type: mRNA
A:Residues: 1-276 <RE2>
A:Cross-references: GB:M57626; NID:g200508; PIDN:AAA39988.1; PID:g200509
R:Reynolds, D.S.; Stevens, R.L.; Lane, W.S.; Carr, M.H.; Austen, K.F.; Serafin, W.E.
Proc. Natl. Acad. Sci. U.S.A. 87, 3230-3234, 1990
A:Title: Different mouse mast cell populations express various combinations of at least
A:Reference number: A38646; MUID:90222202; PMID:2326280

```

Query Match	25.4%;	Score 496;	DB 1;	Length 1019;
Best Local Similarity	39.2%;	Pred. No. 5.6e-31;		
Matches 102;	Conservative	85;	Indels 26;	Gaps 97
QY	38	CGR-----PEPSARIVGSSNAQGTWPMQVSLHHGGHICGSLIAPSWVLSSAAHCFMTWG	93	
Db	772	CGKLLAAQDITPKIVGSSNAKEGAPWVGIVYGRLLCGASLVSSDMLVSSAAHC-VYGR	830	
QY	94	TLPEAAEWSVLGVHSDGDLGHAHT--RAVAATVVPANTYSQVELGADLALLRLASPASL	151	
Db	831	NLEP-SKWTALLGLHMKSN-LTSPQTVPLRDEIVPHNRRKNDIAMHLEFKVNY	888	
QY	152	GPAPVVPCLPRASHRFVHGTCANWAGDV-----OEADPLPLPVLOKVELRLLGEATCQ	207	
Db	889	TDYIQICLPEENQVFPFGRNCSAGTGWVYQGTAN-----ILOADVPLLSNERCQ	942	
QY	208	CLYSOPGFNLTLQILPMLCAGPEGRDRTCQSDSGGPLVCEEGGRWFQAGITSFGPGC	267	
Db	943	---QOMPEYNIT-----ENMICAGYEEGGIDSCQSDSGGPLCMQENNRWFLAGVTSFGYKC	995	
QY	268	GRNRNPGVFTAVATYEAWIR	287	
Db	996	ALPNRPGVIARVSRFTWIO	1015	

RESULT 6

RESULT 6
JC4171
trypsinase (EC 3.4.21.59) precursor - rat
N:Alternate names: mast cell trypsinase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Aug-1995 #sequence_revision 27-Oct-1995 #text_change 20-Jun-2000
C:Accession: JC4171
R:Jide, H.; Itoh, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Maruyama, H.; Osada, Y.; N
J. Biochem. 118, 210-215, 1995
A:Title: cDNA sequencing and expression of rat mast cell trypsinase.
A:Reference number: JC4171; MUID:96015171; PMID:8537314
A:Accession: JC4171
A:Molecule type: mRNA
A:Residues: 1-274 <IDE>
A:Cross-references: DDBJ:D38455; NID:G556555; PIDN:BA07486.1; PID:G556556
C:Comment: This enzyme is basically specific for a connective tissue mast cell, it is up
enzyme inhibitors.
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; hydrolase; mast cell; serine proteinase; zymogen
F:1-15/Domain: signal sequence #status predicted <SIG>
F:20-29/Domain: activation peptide #status predicted <ACT>
F:30-274/Product: mast cell trypsinase #status predicted <MAT>

F:30-266/Domain: trypsin homology <TRY> F:73,120,223/Active site: His, Asp, Ser #status predicted F:131/Binding site: carbohydrate (Asn) (covalent) #status predicted		Query Match 25.2%; Score 491.5; DB 2; Length 274; Best Local Similarity 40.2%; Pred. No. 3e-31; Matches 106; Conservative 28; Mismatches 87; Indels 43; Gaps 6;	
QY	47	IVGGSNAQPGTWQVSLHHGGG---HICGGSLIAPSWLSAAHCFWNTGTLPEAAEWSV	103
Db	30	IVGGRASPKPWQVSLRKFSEFWMHFGGSLIHPQWLTAAHC-----	74
QY	104	LLGVHSQDGLDGAHTR-----AAVALVVPANTYSQVELGADIALRLASPSLG	152
Db	75	-VGLHIKSELPFRVQREQLYADQLLVNRTVHPHYTTVEDGADIALLELPVNV	133
QY	153	PAWVPVCLPRASHRFVHGACWATGVDQVADPLPLPWQLQVEVLRLLGEATCCCLY--	210
Db	134	THIPLSLPPASSETPFGSGTSCWWTGWDIDSDPFLPLPVLPKQVKVPIVENSILCRKXHT	193
QY	211	-----SOPGFENLTQLPGLMCLAGYEGRRDTCOGSGGLPVCEGGRWFQAGITSFGFG	266
Db	194	GLYTGDVPI-----VQDGYLCAG--NTRSDSCQSGDGLPVCKVKGKGTWLGQGVWSWGG	246
QY	267	CGRNRPGVFTAVATYEAWIREQV	290
Db	247	CABANRPGITRYTYLDWLRHV	270
RESULT 7 S00845 hepsin (EC 3.4.21.-) - human C:Species: Homo sapiens (man) C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999 C:Accession: S00845 R:Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W. Biochemistry 27, 1067-1074, 1988 A:title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane dom A:Reference number: S00845; MUID:88209431; PMID:2835076 A:Accession: S00845 A:Molecule type: mRNA A:Residues: 1-417 <LEY> A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064 C:Genetics: A:Gene: GDB:HPN; TMRSS; hepsin A:Cross-references: GDB:133685; OMIM:142440 A:Map position: 19q11-19q13.2 C:Superfamily: hepsin; trypsin homology C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein F:23-45/Domain: transmembrane #status predicted <TMN> F:163-400/Domain: trypsin homology <TRY> F:188-204,291-359,322-338,349-381/Diulfide bonds: #status predicted F:203,257,353/Active site: His, Asp, Ser #status predicted		Query Match 25.1%; Score 490.5; DB 1; Length 417; Best Local Similarity 39.4%; Pred. No. 5.7e-31; Matches 111; Conservative 29; Mismatches 107; Indels 35; Gaps 9;	
QY	38	CGRPE-PSARIVGGSNAQPGTWQVSLHHGGGHCIGGSLIAPSWLSAAHCFWNTGLE	96
Db	153	CGRKLPVDRIVGGRDTSLGRRPWQVSLRYDGAHLGGSLSSGDMVLTAAHCFERNRV-	211
QY	97	PAAEWSVLGVHSDGDLGDAH--TRAVAAVVPANTY-----SOVELGADIALRLASP	148
Db	212	-LSRNRVPAGVAQASP-----HGLQGVQAVVYHGGVLPFRDPNSENSNDIALVHLSPP	266
QY	149	ASLGEPAWVPCVCLPRASHRFVHGACWATGVDV-----QADPLPLPWQLQVEVLRLLGEA	204
Db	267	LPLTEYIQVCLPRAQALVDGKICTVTGWTGNTQYQQAG-----VLQEARVPIISND	320
QY	205	TCQCLYSQGFENLTQLPGLMCLAGYEGRRDTCOGSGGLPVCEEG---GRWFQAGI	260
Db	321	VCN-----GADFYGNQIKPMFCAGYEGGIDACQSGSGGFFVCEDSISRTPRWRLCGI	374
QY		261 TSFGCGGRRNRPGVFTAVATYEAWIREQVVMGSEBPGPAFPQTQ	
Db		375 VSWGTCALAQKPGVYTKVSDFRWIFQAIKTHSEASGMVTQ	
RESULT 8 A35863 trypsin (EC 3.4.21.59) I precursor - human C:Species: Homo sapiens (man) C>Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 21-Jul-2000 C:Accession: A35863; A60939; A39326 R:Vanderslice, P.; Ballinger, S.M.; Tam, B.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G. Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990 A:title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine pr A:Reference number: A35863; MUID:90251647; PMID:2187193 A:Accession: A35863 A:Molecule type: DNA A:Residues: 1-275 <VAN> A:Cross-references: GB:M33494; NID:g3927804; PIDN:AAC83172.1; PID:g3399977 A:Accession: D35863 A:Molecule type: mRNA A:Residues: 1-275 <VA2> A:Cross-references: GB:M33491 R:Butterfield, J.H.; Weiler, D.A.; Hunt, L.W.; Wynn, S.R.; Roche, P.C. J. Leukoc. Biol. 47, 409-419, 1990 A:title: Purification of tryptase from a human mast cell line. A:Reference number: A60939; MUID:90244210; PMID:2110591 A:Accession: A60939 A:Molecule type: protein A:Residues: 31-38, 'P', 40-41, 'X', 43, 'T', 45-48, 'X', 50 <BUT> A:Experimental source: mast cell A:Note: 44-Gly was also found R:Croneish, J.A.; Seidah, N.G.; Marcinkiewicz, M.; Hamelin, J.; Johnson, D.A.; Chretien, J. Biol. Chem. 262, 1363-1373, 1987 A:title: Human pituitary trypsin: molecular forms, NH-2-terminal sequence, immunocytoch A:Reference number: A39326; MUID:87109258; PMID:3543004 A:Accession: A39326 A:Molecule type: protein A:Residues: 31-38 <CRO> A:Experimental source: pituitary C:Genetics: A:Introns: 21/1; 78/2; 177/1; 221/3 A:Superfamily: trypsin; trypsin homology C:Keywords: hydrolase; serine proteinase; zymogen F:1-21/Domain: signal sequence #status predicted <SIG> F:22-30/Domain: activation peptide #status predicted <ACT> F:31-267/Product: trypsin I #status experimental <MAT> F:31-267/Domain: trypsin homology <TRY> F:74,121,224/Active site: His, Asp, Ser #status predicted		Query Match 24.9%; Score 485.5; DB 2; Length 275; Best Local Similarity 41.7%; Pred. No. 9e-31; Matches 111; Conservative 32; Mismatches 106; Indels 17; Gaps 7;	
QY	32	ARGPPVCGPEPARIVGGSNAQPGTWQVSLH-HG--GGHICGSLIAPSWLSAAHC	88
Db	16	AYAAPAGQALQKVGIVGGQEAPEPSKPMQVSLRVHGVPMHFFCGSLIHPQWLTAAHC	75
QY	89	FWNTGTLPEAEWSVLL--GVHSQDGLDGAHTRAVAAVVPANTYSQVELGADIALRL	145
Db	76	--VGPDVKDLAALRVQLREQLYYQQQL-----PVSRIIVHPQFVTAQIGADIALLEL	127
QY	146	ASPAISGPVWPVCLPRASHRFVHGACWATGVDQVADPLPLPWQLQVEVLRLLGEAT	205
Db	128	EFPNVSSVHTVTLPPASSETFPFGMECVTGWGVDNDRLEPPFPFLKQVKVPIMENHI	187
QY	206	CQCLYSQGFENLTQLPGLMCLAGYEGRRDTCOGSGGLPVCEEGGRWFQAGITSFG	264
Db	188	CDARYHLGAYTGDVIRVDDMLCAG--NTRRSDSCQSGDGLPVCKVKGKGTWLGQGVWSG	245
QY	265	FGCGRNRPGVFTAVATYEAWIREQV	290

Db 246 EGCAQPNRPGIYTRVYLDWIHHYV 271

RESULT 9

B35863
N:Alternate names: trypsin - human
C:Species: Homo sapiens (man)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
C:Accession: B35863, A37193, F59473
R:Vanderslice, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990
A:Title: Human mast cell trypsinase: multiple cDNAs and genes reveal a multigene serine protease
A:Reference number: B35863; MUID:90251647; PMID:2187193
A:Accession: B35863
A:Molecule type: mRNA; DNA
A:Residues: 1-275 <VAN>
A:Cross-references: GB:M33492; NID:g339982; PIDN:AAA36779.1; PID:g339983
A:Note: residues 2-275 are derived from mRNA; residue one was inferred from the genomic
R:Miller, J.S.; Moxley, G.; Schwartz, L.B.
J. Clin. Invest. 86, 864-870, 1990
A:Title: Cloning and characterization of a second complementary DNA for human trypsinase.
A:Reference number: A37193; MUID:90369005; PMID:2203827
A:Accession: A37193
A:Molecule type: mRNA
A:Residues: 1-275 <MIL>
A:Cross-references: GB:M37488; NID:g179583; PIDN:AAA51843.1; PID:g179584
R:Blom, T.; Hellman, L.
Scand. J. Immunol. 37, 203-208, 1993
A:Title: Characterization of a trypsinase mRNA expressed in the human basophil cell line K562
A:Reference number: F59473; MUID:93166209; PMID:8434231
A:Accession: F59473
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-275 <RES>
A:Cross-references: GB:S5551; NID:g265666; PIDN:AAD13876.1; PID:g4261576
A:Experimental source: basophil cell line K0812
C:Genetics:
A:Gene: GDB:TPS1
A:Cross-references: GDB:125890; OMIM:191080
A:Map position: lepton-16qter
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-30/Domain: activation peptide #status predicted <ACT>
F:31-275/Domain: trypsin homology <TRY>
F:31-275/Domain: trypsin homology <TRY>
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 24.8%; Score 484.5; DB 2; Length 275;
Best Local Similarity 41.7%; Pred. No. 1.1e-30;
Matches 111; Conservative 31; Mismatches 107; Indels 17; Gaps 7;
Qy 32 ARGPPYGRPPSARIYVGSNAQGTWPQVSLH-HG--GGHICGSLIAPSWLSSAHC 88
Db 16 AYAAPAPQALQRYGVIGQGEAPRSKMPQVSLRVHGFYNNHFCGSLIHPQWLTAAHC 75
Qy 89 FMTNGTLEPAEWSVLG--GVHSDGDLGCAHTRAAAVVVPANYSQVELGADLALLRL 145
Db 76 --VGPDKDLAALRVOLREQLHYDQDL-----PVSRIVHPQFYTAQIGADLALLEL 127
Qy 146 ASPASLGPANVPVCLPRASHRFVHGTCACWATGWDVQVQADPLPLPWLQVYELRLGEAT 205
Db 128 EEPVKVSHVHTVTLPPASETFPPGMPQWVTGWDVNDERLPPFPPLKQVPMENHI 187
Qy 206 CQCLYSQGPFLNLTQIL-PCMLCAGYEGEGRDTCQSGGGLPCEEGRWFGAGITSPG 264
Db 188 CDKATHLGAYTGDDVRIYRDMCLAG--NTRDSCQSGGGLPCKVNGTQLQAGVSWG 245
Qy 265 FCGGRNRPVGFATVATYEAWIREQV 290
Db 246 EGCAQPNRPGIYTRVYLDWIHHYV 271

RESULT 10

KOHUP

plasma kallikrein (EC 3.4.21.34) precursor - human
N:Alternate names: kininogenin; plasma prekallikrein
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
C:Accession: A00921; A37939
R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four trypsin-like units
A:Reference number: A00921; MUID:86243359; PMID:3521732
A:Accession: A00921
A:Molecule type: mRNA
A:Residues: 1-638 <CHU>
A:Cross-references: GB:M13143; NID:g190262; PIDN:AAA60153.1; PID:g190263
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of 1
A:Reference number: A37939; MUID:91152016; PMID:1998666
A:Accession: A37939
A:Molecule type: protein
A:Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;141-1
;260-283,'X',285;287-291,'X',293-295;314-317,'X',319-320;321-324,'X',329-333;334-339,'X',
525;538-551;562,'X',564-567;573,'X',575-576;578-583,'X',585;592-604 <WCM>
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a lig
are linked by one or more disulfide bonds.
C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal re
inogen and may also play a role in the renin-angiotensin system by converting prorenin in
C:Genetics:
A:Gene: GDB:KLK3
A:Cross-references: GDB:127575; OMIM:229000
A:Map position: 4q35-4q35
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inflan
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-638/Product: plasma kallikrein #status predicted <MAT>
F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-621/Domain: trypsin homology <TRY>
F:321-104;47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,383-
F:127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:318-347,340-345/Disulfide bonds: #status predicted
F:330-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 24.7%; Score 482.5; DB 1; Length 638;
Best Local Similarity 40.2%; Pred. No. 3.9e-30;
Matches 104; Conservative 33; Mismatches 97; Indels 25; Gaps 6;
Qy 44 SARIYVGSNAQGTWPQVSLH---HGGHICGSLIAPSWLSSAHCFTWNGTLEPAAE 100
Db 388 STRIVGTNSWGEWPPQVSLQVLTQAQRLCGSLIGHQWLTAAHCF---DGLPLQDV 444
Qy 101 WSVLLGV-----HSQDGLDGAHTRAAVAIVVPANYSQVELGADLALLRLASPASLGPV 155
Db 445 WRIYSGILNLSDTKOTFFS-----QIKELIHHONYKVSQGNHDIALLKQAPLNYTEFQ 499
Qy 156 WPVCLPRASHRFVHGTCACWATGWDVQVQADPLPLPWLQVYELRLGEATCQCLYSQGP 215
Db 500 KPICLPKSGDSTIYTCWVTGWSKEGKIQN--ILQKVNIPLVNTEECQKRYQD--- 554
Qy 216 FNLTQLPLPMLCAGYEGEGRDTCQSGGGLPCEEGRWFGAGITSPGCGRRNPGV 275
Db 555 ----YKLTQRMVCAVYEGEGKQACKGDSGGGLPCKVNGTQLQAGVSWG 610
Qy 276 FTAVATYEAWIREQVMSGE 294

Db 611 YTKVAEYMDWILEKXTQSSD 529

RESULT 11

A45754

trypsinase (EC 3.4.21.59) alpha precursor - human

C:Species: Homo sapiens (man)

C>Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 08-Sep-1997

C:Accession: A45754; B37193

R:Miller, J.S.; Westin, E.H.; Schwartz, L.B.

J. Clin. Invest. 84, 1188-1195, 1989

A:Title: Cloning and characterization of complementary DNA for human trypsinase.

A:Reference number: A45754; MUID:90009311; PMID:2677049

A:Accession: A45754

A:Molecule type: mRNA

A:Residues: 1-274 <MIL>

A:Cross-references: GB:M30038

R:Miller, J.S.; Moxley, G.; Schwartz, L.B.

J. Clin. Invest. 86, 864-870, 1990

A:Title: Cloning and characterization of a second complementary DNA for human trypsinase.

A:Reference number: A37193; MUID:90369005; PMID:2203827

A:Accession: B37193

A:Molecule type: mRNA

A:Residues: 1-274 <M12>

A:Cross-references: GB:M30038

A>Note: the sequence from Fig. 4 is inconsistent with that from Fig. 2 in having 205-Pro

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase; zymogen

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-30/Domain: activation peptide #status predicted <ACT>

F:31-274/Product: trypsinase I #status predicted <MAT>

F:31-266/Domain: trypsin homology <TRY>

F:74,120,223/Active site: His, Asp, Ser #status predicted

Query Match 24.7%; Score 481.5; DB 2; Length 274;

Best Local Similarity 40.8%; Pred. No. 1.8e-30;

Matches 108; Conservative 31; Mismatches 97; Indels 29; Gaps 7;

QY 41 PEP-----SARIVGGNAQPGTWPQVSLH--HGGHICGSLIAPSWSLVAHCFMIN 92

Db 20 PAPVQALQAGIVGGQAPRKPQVSLRVRDRYWHFGGSLHPQWVLIHANC--- 75

QY 93 GTLEPAEWSVLGVHSQDGLDGAH-----TRAVAAIVVPANYSQVELGADLALLRLA 146

Db 76 --LGPDKOLATRVNS-----GTHLYYQDQLLPVSRIMVHPQFVLIQGDADIALLELE 127

QY 147 SPASLGPVWPVCLPRASHRFVHGTCATWGVDVQADPLPLPWLQVEVELLGEATC 206

Db 128 EPVNISSRVHTVMLPPASSETFPFGMPCWVTGWGDVNDDELPFPFLKQVKVFMENHIC 187

QY 207 QCLYSQPGPNLTILQIL-PCMLCAGYPEGRDTCQDGGPLVCEEGRWFOAGITISFGF 265

Db 188 DAKTHLGAYTGDDVKRIIRDMCLAG--NSQRDCKGDSGGPLVCKVNGTWLQAGVSWDE 245

QY 266 GCGRRNRPVGTAVATYEAIRREQV 290

Db 246 GCAQNPNGIVTRYTYLDLMIHV 270

RESULT 12

A53663

enteropeptidase (EC 3.4.21.9) precursor [validated] - pig

N:Alternate names: enterokinase

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003

C:Accession: A53663

R:Matsumura, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa, J. Biol. Chem. 269, 19976-19982, 1994

A:Title: Structural characterization of porcine enteropeptidase.

A:Reference number: A53663; MUID:94327548; PMID:8051081

A:Accession: A53663

A:Molecule type: mRNA

A:Residues: 1-1034 <MAT>

A:Cross-references: GB:D30799; NID:G505122; PIDN:BAA06459.1; PID:G505123

A>Note: parts of this sequence, including the amino ends of three chains isolated from

C:Comment: The mechanism of association with the membrane of the intestinal brush border

ated below) or with amino-terminal myristoylation of the heavy chain.

C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)

ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv

C:Function:

A:Description: cleaves activation peptide from trypsinogen to produce active trypsin

A:Pathway: intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; Ctr/Cls repeat homology; LDL receptor ligand-binding ref

C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen

F:22-38/Domain: transmembrane #status predicted <TM>

F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>

F:118-799/Product: enteropeptidase heavy chain #status predicted <HCH>

F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:357-519/Domain: MAM homology <MAM>

F:541-646/Domain: Ctr/Cls repeat homology <CLR>

F:658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC

F:800-1034/Product: enteropeptidase light chain #status predicted <LCH>

F:800-1029/Domain: trypsin homology <TRY>

F:116,147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,96

F:787-911,825-841,925-992,956-971,982-1010/Disulfide bonds: #status predicted

F:840,891,986/Active site: His, Asp, Ser #status predicted

Query Match 24.6%; Score 480; DB 1; Length 1034;

Best Local Similarity 38.5%; Pred. No. 1e-29;

Matches 99; Conservative 46; Mismatches 92; Indels 20; Gaps 9;

QY 38 CGR-----PEPSARIVGGNAQPGTWPQVSLHGGHICGSLIAPSWSLVAHCFMUNG 93

Db 787 CGKQVQAEVSPKIVGGNDSREGAWPVVVALYNGQLICGASLVSRDLWLSAHC-VYGR 845

QY 94 TLEPAEWSVLGVHSQDGLDGAH--TRAVAAIVVPANYSQVELGADLALLRLASPASL 151

Db 846 NLEP-SKWKAILGLEMTSN-LTSPQIVTLRIDEIVINPHYNNRRKDSOTAMHLEFKVNY 903

QY 152 GPVAPVWPVCLPRASHRFVHGTCATWGVDV-QEADPLPLPWLQVEVELLGEATCQCCLY 210

Db 904 TDVITQICLPEENQVPPGRCISAGWGKVIYQGGSPAD---ILOEADVELLSNEKQ--- 957

QY 211 SQPGPNLTILQIL-PCMLCAGYPEGRDTCQDGGPLVCEEGRWFOAGITISFGCGRR 270

Db 958 QQMEYNIT-----ENWMCAGYEGGIDSCQDGGPLVCEEGRWFOAGITISFGYQCALP 1013

QY 271 NRPVGTAVATYEAIR 287

Db 1014 NRPVGTAVATYEAIR 1030

RESULT 13

C35863

trypsinase (EC 3.4.21.59) III precursor - human

C:Species: Homo sapiens (man)

C>Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-Mar-2003

C:Accession: C35863; E35863; A38893

R:Vanderslice, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.

proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990

A:Title: Human mast cell trypsinase: multiple cDNAs and genes reveal a multigene serine pr

A:Reference number: A35863; MUID:90251647; PMID:2187193

A:Accession: C35863

A:Molecule type: mRNA

A:Residues: 9-275 <VAN>

A:Accession: E35863

A:Molecule type: DNA

A:Residues: 1-9 <VA2>

A:Cross-references: GB:M33494; NID:G3927804; PIDN:AA083172.1; PID:G339977

A>Note: the first nine residues of this sequence are inferred from genomic DNA of trypte

R:Vanderslice, P.

submitted to GenBank, April 1990

A:Reference number: A38893

A:Accession: A38893

A:Molecule type: mRNA

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 14:37:37 ; Search time 49 Seconds
(without alignments)
379.368 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKGVLPQGLGAVANSDS.....TKSLVLFWLSPHSLGLGWGF 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	690	35.3	343	1 PSS8 HUMAN	Q16651 homo sapien
2	631	32.3	342	1 PSS8 MOUSE	Q9esd1 mus musculu
3	620	31.7	342	1 PSS8 RAT	Q9es87 rattus norv
4	582	29.8	321	1 TRYX HUMAN	Q9nr22 homo sapien
5	563	28.8	290	1 PR27 HUMAN	Q9bqr3 homo sapien
6	547	28.0	811	1 TMS6 HUMAN	Q8iu80 homo sapien
7	544	27.9	317	1 BSS4 HUMAN	Q8gzd4 homo sapien
8	540	27.6	855	1 ST14 HUMAN	Q9y5y6 homo sapien
9	534.5	27.4	314	1 TEST HUMAN	Q9y6m0 homo sapien
10	533	27.3	811	1 TMS6 MOUSE	Q9db10 mus musculu
11	529	27.1	324	1 TEST MOUSE	Q9jhj7 mus musculu
12	528.5	27.1	311	1 TRYX MOUSE	Q9qlu7 mus musculu
13	523	26.8	855	1 ST14 MOUSE	P56677 mus musculu
14	516	26.4	306	1 BSS4 MOUSE	Q9er10 mus musculu
15	514.5	26.3	275	1 TRYX CANFA	P15944 canis famil
16	508.5	26.0	273	1 TRYX SHEEP	Q9xsm2 ovis aries
17	499	25.6	276	1 TRYX PIG	Q9n2d1 sus scrofa
18	499	25.6	276	1 MCT6 MOUSE	P21845 mus musculu
19	496	25.4	1019	1 ENTX HUMAN	P98073 homo sapien
20	494.5	25.3	454	1 TMS3 HUMAN	P57727 homo sapien
21	494	25.3	453	1 TMS3 MOUSE	Q8k1t0 mus musculu
22	491.5	25.2	274	1 MCT6 RAT	P50343 rattus norv
23	490.5	25.1	417	1 HEPS HUMAN	P05981 homo sapien
24	490.5	25.1	436	1 HEPS MOUSE	Q35453 mus musculu
25	485.5	24.9	275	1 TRB1 HUMAN	Q15661 homo sapien
26	484.5	24.8	275	1 TRB2 HUMAN	P20231 homo sapien
27	482.5	24.7	638	1 KAL HUMAN	P01952 homo sapien
28	480	24.6	1034	1 ENTX PIG	P98074 sus scrofa
29	479	24.5	275	1 TRYX HUMAN	P15157 homo sapien
30	478.5	24.5	416	1 HEPS RAT	Q95511 rattus norv
31	477.5	24.4	457	1 TMS5 HUMAN	Q9h3s3 homo sapien
32	475	24.4	638	1 KAL MOUSE	P36262 mus musculu
33	474.5	24.3	270	1 TRYX MERON	P50342 meriones un

RESULT 1
PSS8 HUMAN
ID PSS8 HUMAN STANDARD; PRT; 343 AA.
AC Q16651; Q9UCA3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Proctasin precursor (EC 3.4.21.-).
GN PSS8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Prostate;
RX MEDLINE=95286644; PubMed=7768952;
RA Yu J.X.; Chao L.; Chao J.;
RT "Molecular cloning, tissue-specific expression, and cellular
localization of human proctasin mRNA."
RL J. Biol. Chem. 270:13483-13489(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullah S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko I., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 45-64.
RC TISSUE=Semen;
RX MEDLINE=94308140; PubMed=8034638;
RA Yu J.X.; Chao L.; Chao J.;
RT "Proctasin is a novel human serine proteinase from seminal fluid.
Purification, tissue distribution, and localization in prostate
gland."
RL J. Biol. Chem. 269:18843-18848(1994).
CC -I- FUNCTION: Possesses a trypsin-like cleavage specificity.
CC -I- SUBUNIT: Heterodimer of two chains, light and heavy, held by a

ALIGNMENTS

FT	PROPEP	323	342	BY SIMILARITY.
FT	TRANSMEM	320	340	POTENTIAL.
FT	DOMAIN	45	286	SERINE PROTEASE.
FT	DISULFID	37	154	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	70	86	BY SIMILARITY.
FT	DISULFID	168	244	BY SIMILARITY.
FT	DISULFID	201	223	BY SIMILARITY.
FT	DISULFID	234	262	BY SIMILARITY.
FT	ACT_SITE	85	85	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	134	134	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	238	238	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CARBOHYD	110	110	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	159	159	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	342 AA;	36729 MW;	0620DE88ED187D0F CRC64;
Query Match 32.3%; Score 631; DB 1; Length 342;				
Best Local Similarity 41.4%; Pred. No. 5.2e-43;				
Matches 146; Conservative 48; Mismatches 139; Indels 20; Gaps 10;				
QY	1	MAQKGVLPQQLGAVANS	DSYSLYGLVPSG-PARG-PPYCGRP-EP	SARIVGGSSNAQPGT 57
Db	1	MAPRVGLGLQLEAVT---	ILLGLLQSGIRADGTEASCGAVIQP--	RITGGGSAKPGQ 55
QY	58	WPQVSLHGGGHI	CGGSLIAPSWLSAAHCFMTNGTLEPAAE	WSVLLGVHSDQDPLDGA 117
Db	56	WPQVSLTYDGNHVC	GSLSVSNKVVSAAHCFPREHSRE---	AYEVKLGALHQLDSYSNDT 112
QY	118	HTRAVAAIVPANY	SQVELGADLALLRLASPASLGPAVWPVCL	PRASHRFVHGTACWATG 177
Db	113	VVHTVAQIITHSSY	REGSGQDIAFIRLSSPVTFSSYIRPICL	PAANAFNGLHCTVTG 172
QY	178	WGDVQADPLPLPW	LQVVELRLLEATCCQLYSQPGFNNLTQIL	PGMLCAGYPEGRRD 237
Db	173	WGHVAPSVLSQTP	RLPQLQLEVLISRETCSCLYNINAVPEE	PHTIQDMLCAGYVKGKD 232
QY	238	TCQGSGLPLCBEGR	WFOAGITSPFCGGRNRPGVFTAVATYPA	WIREQVMGSEPGP 297
Db	233	ACQGSGLPLSCPME	GIWYLAGIVSGDACGAPNRPGVYTLTST	YASWIHHV--AELQP 290
QY	298	AFPTQPKTKQSD---	CLHQTAFLDSAR--ILRLPLSHISVGVST	GTSKSLVPLWL 346
Db	291	RVVPQTQESQPDG	HLCHNHPVFNAAAPKLLRPLFLPLGLTG	LLSL---WL 340
RESULT 3				
PSS8 RAT STANDARD; PRT; 342 AA.				
AC	Q9ESB7; Q9ER01;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Prostasin precursor (EC 3.4.21.-).			
GN	PKS8.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	Tissue=Kidney;			
RA	Adachi M., Kitamura K., Miyoshi T., Tomita K.;			
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Wang C.;			
RT	"Molecular cloning and expression of rat prostasin.";			
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: Possesses a trypsin-like cleavage specificity (By			
CC	similarity).			
CC	-!- SUBUNIT: Heterodimer of two chains, light and heavy, held by a			
CC	disulfide bond (By similarity).			
CC	-!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF			
CC	ITS C-TERMINUS (BY SIMILARITY).			

CC	-!- SIMILARITY: Belongs to peptidase family S1.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AB017638; BAB20281.1; --			
DR	EMBL; AF202076; AAG32641.1; --			
DR	HSSP; P00734; IUVS.			
DR	InterPro; IPR009003; Cys_Ser_trypsin.			
DR	InterPro; IPR001254; Peptidase_S1.			
DR	InterPro; IPR001314; Peptidase_S1A.			
DR	Pfam; PF00089; trypsin; 1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00020; Tryp_Spc; 1.			
DR	PROSITE; PS50240; TRYPSIN_DOM; 1.			
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.			
DR	PROSITE; PS00135; TRYPSIN_SER; 1.			
KW	Hydrolase; Serine protease; Zymogen; Signal; Glycoprotein;			
KW	Transmembrane.			
FT	SIGNAL	1	29	POTENTIAL.
FT	PROPEP	30	32	ACTIVATION PEPTIDE (BY SIMILARITY).
FT	CHAIN	33	44	PROSTATIN LIGHT CHAIN.
FT	CHAIN	45	322	PROSTATIN HEAVY CHAIN.
FT	PROPEP	323	342	BY SIMILARITY.
FT	TRANSMEM	320	340	POTENTIAL.
FT	DOMAIN	45	286	SERINE PROTEASE.
FT	DISULFID	37	154	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	70	86	BY SIMILARITY.
FT	DISULFID	168	244	BY SIMILARITY.
FT	DISULFID	201	223	BY SIMILARITY.
FT	DISULFID	234	262	BY SIMILARITY.
FT	ACT_SITE	85	85	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	134	134	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	238	238	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CARBOHYD	159	159	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	17	17	I -> V (IN REF. 1).
FT	CONFLICT	292	292	A -> V (IN REF. 1).
SQ	SEQUENCE	342 AA;	36843 MW;	5EDIAF05D9213B98 CRC64;
Query Match 31.7%; Score 620; DB 1; Length 342;				
Best Local Similarity 40.8%; Pred. No. 3.9e-42;				
Matches 144; Conservative 50; Mismatches 139; Indels 20; Gaps 9;				
QY	1	MAQKVGPGQLGAVANS	DSYSLYGLVPS--GPARGPPYCGRP-BPSARI	VGGSSNAQPGT 57
Db	1	MAPRVGLGQLEALF---	ILLTGLLQSGIRADGTEASCGAVIQP--RITGGGSAKPGQ 55	
QY	58	WPQVSLRHGGGHI	CGGSLIAPSWLSAAHCFMTNGTLEPAAEWSVLLGVHSDQDGLDGA 117	
Db	56	WPQVSLTYDGNHVC	GSLSVSNQVWSAAHCFPREHSKE---EYEVKLGALHQLDSFSNDI 112	
QY	118	HTRAVAAIVVPANY	SQVELGADLALLRLASPASLGPAVWPVCLPRASHRFVHGTACWATG 177	
Db	113	VVHTVAQIITHSSY	REGSGQDIALIRLSSPVTFSSYIRPICLPAANAFPNGLHCTVTG 172	
QY	178	WGDVQEADPLPLPW	LQVQVELRLLEATCCQLYSQPGPNLTLQILPGMLCAGYPEGRED 237	
Db	173	WGHVAPSVLSQTP	RLPRLQQLLEVLISRETCSCLYNINAVPEEPTIQDMLCAGYVKGKD 232	
QY	238	TCQGSGLPLCBEGR	WFOAGITSPFCGGRNRPGVFTAVATYPAWIREQVMGSEPGP 297	
Db	233	ACQGSGLPLSCPD	GLWYLAGIVSGDAGCAPNRPGVYTLTSTYASWIHHV--AELQP 290	
QY	298	AFPTQPKTKQSD---	CLHQTAFL-IDSARILLRLPLSHISVGVSTGTSKSLVPLWL 346	
Db	291	RAVPQTQESQPDG	HLCHNHPVFNIAAOKSRPILFLPLSLTGLFSL--WL 340	


```

RESULT 4
TRYG HUMAN
ID TRYG HUMAN STANDARD; PRT; 321 AA.
AC Q9NR2; Q9C015; Q9NR08; Q9UBB2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Trypsin gamma precursor (EC 3.4.21.-) (Transmembrane trypsinase).
GN TP5G1 OR TMT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RX MEDLINE=20302813; PubMed=10843716;
RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallao M.,
RA Wolters P.J., Verghese G.M.;
RT "Characterization of human gamma-tryptases, novel members of the
RT chromosome 16p mast cell trypsin and prostatic gene families.";
RT J. Immunol. 164:6566-6575 (2000).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99452974; PubMed=10521469;
RA Wong G.W., Tang Y., Feyfant E., Sali A., Li L., Li Y., Huang C.,
RA Friend D.S., Krill S.A., Stevens R.L.; trypsinase family of mouse and
RT human mast cell proteases which possess a novel COOH-terminal
RT hydrophobic extension.";
RT J. Biol. Chem. 274:30784-30793 (1999).
RN [3]
RN SEQUENCE OF 220-321 FROM N.A.
RA Mittman S., Agnew W.S.;
RT "Organization and alternative splicing of CACNA1H.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Membrane-anchored (Potential).
CC -!- TISSUE SPECIFICITY: Expressed in many tissues.
CC -!- POLYMORPHISM: There are two alleles; gamma-I and gamma-II which
CC differ by 5 residues.
CC -!- SIMILARITY: Belongs to peptidase family S1. Trypsin subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF191031; AAF76457.1; -
CC EMBL; AF195508; AAF76458.1; -
CC EMBL; AF175759; AAF03697.1; -
CC EMBL; AF175522; AAF03695.1; -
CC EMBL; AF223563; AAG48852.2; -
CC HSP; P00763; 1DPO.
CC MEROPS; S01.028; -.
CC Genew; HGNC:14134; TP5G1.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR01254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00089; trypsin, 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PROSITE; PS02240; TRYPsin DOM; 1.
CC PROSITE; PS00134; TRYPsin_HIS; 1.
CC PROSITE; PS00135; TRYPsin_SER; FALSE NEG.
CC KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;
KW Transmembrane; Polymorphism.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 36 TRYPTASE GAMMA LIGHT CHAIN.
FT CHAIN 38 321 TRYPTASE GAMMA HEAVY CHAIN.
FT TRANSMEM 284 304 POTENTIAL.
FT ACT_SITE 78 78 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 125 125 CHARGE RELAY SYSTEM (BY SIMILARITY).

```

```

FT ACT_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 26 145 INTERCHAIN (POTENTIAL).
FT DISULFID 63 79 BY SIMILARITY.
FT DISULFID 159 228 BY SIMILARITY.
FT DISULFID 192 210 BY SIMILARITY.
FT DISULFID 218 246 BY SIMILARITY.
FT CARBOHYD 85 85 M-LINKED (GLCNAC..) (POTENTIAL).
FT VARIANT 60 60 M -> V (IN GAMMA-II).
FT VARIANT 126 126 I -> M (IN GAMMA-II).
FT VARIANT 132 132 S -> T (IN GAMMA-II).
FT VARIANT 204 204 L -> I (IN GAMMA-II).
FT VARIANT 288 288 L -> F (IN GAMMA-II).
FT CONFLICT 160 160 W -> S (IN REF. 1).
SQ SEQUENCE 321 AA; 33827 MW; FFF7B0E3C4A962D CRC64;

Query Match 29.8%; Score 582; DB 1; Length 321;
Best Local Similarity 46.0%; Pred. No. 3.7e-39;
Matches 128; Conservative 35; Mismatches 87; Indels 28; Gaps 11;

QY 36 PYCGRPEPS---ARIYVGSNAQPGTPWQVSIHHGGHICGSLIAPSWVLSAAHCFWTFN 92
Db 24 PCGGRPQVSDAGRIYVGGHAPAGAPWQASIRLRMRHVCGSLLSPQWLVTAHCF--S 81
QY 93 GTLEPAASVLLGVHSGDGLD---GAHTRAVAAIVVPANYS-QVELGADIALRLASP 148
Db 82 GSLN-SSDYQVHL-----GELEITLSPHFSTVROIHSFSGQPGTSGDIALVELSVP 134
QY 149 ASLGPVAVPVCILPRASHRFVHGTACWATGWGDVQVADLPPLPWLOFVELRLGEATQC 208
Db 135 VTLSRIILPVCLPEASDDFCPIRCWVIGWYREGEPLPEYSLSREKVSVDVETCR 194
QY 209 LYSQPGPNLTLQIIPGMLCAGYPERDRDTCQDGGPLVCEGGRWFQAGITSFGCG 268
Db 195 DY--FPGGSIQ--FDMLCA---RPGDGCDDSGPLVCQVNGAWVQAGIVSWGCGG 247
QY 269 RNRPGVTAATYEAATREQVW---GSEPGPAFTQP 303
Db 248 RNRPGVTRPAYVWVNRHITASGSESG--YPLP 283

RESULT 5
PR27_HUMAN STANDARD; PRT; 290 AA.
AC Q9BQR3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Pancreas precursor (EC 3.4.21.-) (Marapsin) (Channel-activating
DE protease 2) (CAPH2).
GN PRSS27 OR MEN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RA Fortunato M., Dando P.M., Rawlings N.D., Barrett A.J.;
RT "Cloning, sequencing and expression of marapsin, a human serine
RT proteinase.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RA Okaze H., Hayashi A., Kozuma S., Saito T.;
RA Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RA MEDLINE=22439795; PubMed=12441343;
RA Bhagwandin V.J., Hau L.W.-T., Mallen-St Clair J., Wolters P.J.,

```

Caughey G.H.;
 "Structure and activity of human pancreas, a novel tryptic serine
 peptidase expressed primarily by the pancreas.";
 J. Biol. Chem. 278:3363-3371(2003).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Expressed predominantly in the pancreas.
 CC -1- PTM: N-glycosylated.
 CC -1- SIMILARITY: Belongs to peptidase family S1.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 or send an email to license@sib-sib.ch.

 DR EMBL; AJ306593; CAC35467.1; -
 DR EMBL; AB056161; BAB85497.1; -
 DR EMBL; AY030095; AAX38168.1; -
 DR HSSP; P00734; 1UVS.
 DR MEROPS; S01.074; -
 DR Genes; HGNC:15475; PRS27.
 DR MIM; 608018; -
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS00240; TRYPsin DOM; 1.
 DR PROSITE; PS00134; TRYPsin HIS; 1.
 DR PROSITE; PS00135; TRYPsin SER; 1.
 DR Hydrolase; Serine protease; Zymogen; Signal; Glycoprotein.
 FT SIGNAL 1 22 POTENTIAL.
 FT PROPEP 23 34 POTENTIAL.
 FT CHAIN 35 290 PANCREASIN.
 FT DOMAIN 35 277 SERINE PROTEASE.
 FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 124 124 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 229 229 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 60 76 BY SIMILARITY.
 FT DISULFID 158 235 BY SIMILARITY.
 FT DISULFID 191 214 BY SIMILARITY.
 FT DISULFID 225 253 BY SIMILARITY.
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 290 AA; 31940 MW; 67BDC93EC70BFF7B CRC64;
 Query Match 28.8%; Score 563; DB 1; Length 290;
 Best Local Similarity 42.2%; Pred. No. 1.1e-37;
 Matches 117; Conservative 32; Mismatches 114; Indels 14; Gaps 6;
 QY 32 ARGPPYCRPEARIYVGSNAQPTWQVSLHGGGHICGSLIAPSWVLSAAHCFMT 91
 Db 20 AKATACGRPMNLNRMVGGQDQEGEWPMQVSIQRNGSHFCGSLIACQWVLTAAHCF-- 77
 QY 92 NGTLEPAAEWSVLLGVHSGDGLDGAHTRAAVAVVPAN--YSQVELGADLALLFLASPA 149
 Db 78 NNTSETSL-YQVLLGARQLVQP--GPHAMVYRVQVSNFLYQGTASSADVALVEAPV 134
 QY 150 SIGPAAWVPVCLPRASHRFVHGATCATWGVDVQVQVADPLPLPWVQVVELLLGEATCOCL 209
 Db 135 PFTNYLPLVCLPDPSPVIFETGMNVCWVTGWGSPSEEDLPEPRILQKLAVPITDTPKCNLL 194
 QY 210 YSQGPEFNLTQLLPQ-MLCAGYVGRDTCOGDGGPLVCEGGWFGWFOAGITTSFGGG 268
 Db 195 YSKOTFEGYQPTIKNDMLCAGFEKDKDCKGSGPLVCLVGSWLGQVISMWGCA 254
 QY 269 RNRNPGVATVATYEAWIREQVWGSPPGPAFTQPOK 305
 Db 255 RNRNPGVYRTVAHNWIHRII-----PKLQFQPAR 285

RESULT 6
 TMS6 HUMAN STANDARD; PRT; 811 AA.
 ID Q8IU80; Q8IXV8;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transmembrane protease, serine 6 (EC 3.4.21.-) (Matrptase-2).
 GN TMPRSS6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Hooper J.D., Quigley J.P.;
 RT "TMPRSS6, a new type II transmembrane serine protease.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Begguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudon J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Mink P.,
 RA Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
 RA Budarf M.I., Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard M., Kiedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Walkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tilahun Y., Wright H.;
 RT "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain.
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Mitsu S., Okui A., Kominami K., Yamaguchi N.;
 RT "Cloning and characterization of a human brain-specific serine
 protease, hBSP-4.";
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Wong G.W., Stevens R.L.;
 RT "Identification of a new member of the chromosome 16 family of serine
 proteases.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McWay P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]
 RP SEQUENCE OF 47-317 FROM N.A.
 RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
 RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
 RA Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,
 RA Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC
 CC EMBL; AB010779; BAB20263.1; -;
 CC EMBL; AF321182; AAG35070.1; -;
 CC EMBL; BC009726; AAB93671.1; -;
 CC EMBL; AC003965; AAB93671.1; -;
 CC HSSP; P00763; 1DPO.
 CC MEROPS; S01.252; -;
 CC Genew; HGNC:14368; PRSS22.
 CC InterPro; IPR009003; Cys_Ser_trypsin.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001314; Peptidase_S1A.
 CC Pfam; PF00089; trypsin; 1.
 CC PRINTS; PR00722; CHYMOTRYPSIN.
 CC SMART; SM00020; Tryp_SPC; 1.
 CC PROSITE; PS50240; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Signal.
 FT SIGNAL 1 32 POTENTIAL.
 FT CHAIN 33 317 BRAIN-SPECIFIC SERINE PROTEASE 4.
 FT ACT_SITE 90 90 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 242 242 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 75 91 BY SIMILARITY.
 FT DISULFID 175 248 BY SIMILARITY.
 FT DISULFID 208 227 BY SIMILARITY.
 FT DISULFID 238 266 BY SIMILARITY.
 FT CARBOHYD 70 70 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 47 47 L -> M (IN REF. 4).
 SQ SEQUENCE 317 AA; 33731 MW; E2A123BC86E79935 CRC64;
 Query Match 27.9%; Score 544; DB 1; Length 317;
 Best Local Similarity 39.3%; Pred. No. 3.8e-36;
 Matches 114; Conservative 41; Mismatches 123; Indels 12; Gaps 4;
 QY 7 LGPQLGAVANSDSYSLYGLVPSGPARGPYCGRPEPSARIVGSGNSAQPTGTPWQVSLHH 66
 DB 10 LGGCLGHTFTSILLASTAILNAARIPVPACGRPQQLNRVVGEDSTDSEWPVLSIQK 69
 QY 67 GGGHICGSLIAPSVLGAHCFMTNGTLEPAEWSVLLGVHSQDGLDGAHTRAVAAIV 126
 DB 70 NGTHRCAGSLTSTRVITAAHCFKDN--LNKPYLFSVLLGAWQLGNPGSRSQKGVAVWE 127
 QY 127 VPANYSQVE--LGADLALRLASPLASGLPAVMPVCLPRASHRFVHGTACWATGWDVQAD 185
 DB 128 PHPVYSWKEGACADIALVRLRSQTSFSEVPICLDPDASIHLPNTHCWISGWSIQDGV 187
 QY 186 PLPLPWLQVEVLELLRLLGRATCQCLY---SQPGPNLTILQLPGLMFCAGYPEGRRDTCCGD 242
 DB 188 PLPHPTQLKLVPIIDSEVCSHLVWRCAGQP-----ITEDMLCAGYLEGERDACLGD 241
 QY 243 SGGPIVCEGGRWFOAGITSPFGCGRRNRPGVFTAVATYEAWIREQVWG 292
 DB 242 SGGPLMCOVDGAWILAGIISWEGCAERNRPGVITSLSAHRSWVEKIVQG 291
 RESULT 8
 ST14 HUMAN
 ID ST14 HUMAN STANDARD; PRT; 855 AA.
 AC Q9YSY6; Q9BS01; Q9H3S0; Q9HB36; Q9HCA3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matriptase) (Membrane-
 DE type serine protease 1) (MT-SPI) (prostatin) (Serine protease TADG-15)
 DE (Tumor associated differentially-expressed gene-15 protein).
 GN ST14 OR PRSS14 OR SNC19 OR TADG15.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99303581; PubMed=10373424;
 RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
 RT "Molecular cloning of cDNA for matriptase, a matrix-degrading serine
 RT protease with trypsin-like activity.";
 RL J. Biol. Chem. 274:18231-18236 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99432178; PubMed=10500122;
 RA Takeuchi T., Shuman M.A., Craik C.S.;
 RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
 RT dissect complex biological processes and identify a membrane-type
 RT serine protease in epithelial cancer and normal tissue.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061 (1999).
 RN [3]
 RP SEQUENCE FROM N.A.

RC TISSUE=Prostate;
RA Yamaguchi N., Mitsui S.;
RT "Molecular cloning of a novel transmembrane serine protease expressed
in human prostate.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Tanimoto H., Underwood L.J., Wang Y., Shigenasa K., Farnley T.H.,
RA O'Brien T.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA TISSUE=Blood and Muscle;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [6]
RP SEQUENCE OF 340-664 FROM N.A.
RA Cao J., Fan W., Zheng S.;
RT "Genomic analysis of a novel human serine protease SNCL9.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=99303582; PubMed=10373425;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matrix
and a kunitz-type serine protease inhibitor from human milk.";
RL J. Biol. Chem. 274:18237-18242 (1999).
CC -!- FUNCTION: Degrades extracellular matrix. Proposed to play a role
in breast cancer invasion and metastasis. Exhibits trypsin-like
activity as defined by cleavage of synthetic substrates with Arg
or Lys as the P1 site.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF118224; AAD42765.2; -
DR EMBL; AF133086; AAF00109.1; -
DR EMBL; AB030036; BAB20376.1; -
DR EMBL; AF057145; AAG15395.1; -
DR EMBL; BC005826; AAH05826.1; -
DR EMBL; BC030532; AAH30532.1; -
DR EMBL; AF283256; AAG13949.1; -
DR HSSP; P00763; IDPO.
DR Genew; HGNC:11344; ST14.

DR MIM; 606797; -
DR MEROES; S01.302; -
DR GO; GO:0005887; C: integral to plasma membrane; TAS.
DR GO; GO:0008236; F: serine-type peptidase activity; TAS.
DR GO; GO:0006508; P: proteolysis and peptidolysis; TAS.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLa; 3.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS50068; LDLRA_2; 4.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
KW Transmembrane; Repeat.
FT DOMAIN 1 55
FT TRANSMEM 56 76
FT DOMAIN 77 855
FT DOMAIN 214 334
FT DOMAIN 340 447
FT DOMAIN 452 487
FT DOMAIN 487 524
FT DOMAIN 524 560
FT DOMAIN 566 603
FT DOMAIN 615 854
FT ACT_SITE 656 656
FT ACT_SITE 711 711
FT ACT_SITE 805 805
FT ACT_SITE 109 109
FT CARBOHYD 302 302
FT CARBOHYD 485 485
FT CARBOHYD 772 772
FT CONFLICT 327 329
FT CONFLICT 381 381
FT CONFLICT 674 674
FT SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;
SQ
Query Match 27.6%; Score 540; DB 1; Length 855;
Best Local Similarity 40.2%; Pred. No. 2.4e-35;
Matches 113; Conservative 39; Mismatches 95; Indels 34; Gaps 7;
QY 32 ARGPPYCGRPEP-----SARIVGGSNAQPGTWPMQVSLHH-GGGHI 71
DB 581 SKGNPECDKDCSDGSDKDCGLRSFTQARVVGTDADGEGWPMQVSLHAGQGGHI 640
QY 72 CGGSLIAPSWLVAHAHCFMTNG--TLEPAEWSVLLGVHSQ-DGPLDGAHTRAVAAIVVP 128
DB 641 CGASLISFNWLSVAHCHYIDDRGFYSPTQTWAFGLHDQSQRSAFGVQERRLKRISH 700
QY 129 ANYSQVEIGADLALRLASPASLGAFFVWVCLPRASHFVHGTCATGWGVDVQADPLP 188
DB 701 PFENDFTDYDIALLELEKPAEYSMWVRPICLPDASHVFPAGKAIWVTGWGHTQGGTGA 760
QY 189 LPWVLOVEVLRLLGEATCCQLYSGPGPENLTILPGMLCAGYPEGRRDTCQDSSGGPL- 247
DB 761 L--ILKGEIRVINGTTCENLLPQ-----QITPRMCMVGFSGGVDSCQDSSGGPLS 810
QY 248 VCEGGGRWFQAGITSFGFGCCRRNRPVGFVTAATVYEAWIRE 288
DB 811 SVEADGRIFQAGVSWGSDGCAQRNKPVGYYTRLPFRDWIKE 851

Query Match 26.4%; Score 516; DB 1; Length 306;
Best Local Similarity 37.1%; Pred. No. 6.2e-34;
Matches 109; Conservative 47; Mismatches 118; Indels 20; Gaps 6;

FT PROPEP 21 30 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 31 275 TRYPTASE.
FT ACT_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 224 224 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 59 75 BY SIMILARITY.
FT DISULFID 155 230 BY SIMILARITY.
FT DISULFID 188 211 BY SIMILARITY.
FT DISULFID 220 248 BY SIMILARITY.
FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 275 AA; 30088 MW; C3869251F248D58 CRC64;

Query Match 26.3%; Score 514.5; DB 1; Length 275;
Best Local Similarity 42.2%; Pred. No. 7.2e-34;
Matches 124; Conservative 27; Mismatches 100; Indels 43; Gaps 9;

QY 23 IYGLVPSGPARGPPYCGRPEPSARIYVCGSNAQPTWPMQVSLHHGG---GHICGGSLLAP 79
Db 12 LGSLLVPVSPAP-----GQALQRVGIVGREGAPGSKWPMQVSLRLKGQYWRHICGGSLLHP 66

QY 80 SWVLSAAHCFMTNGTLEPAAEWSVLL---GVHSQDGLDGAHTRAVAAIIVPEANYSQVEL 136
Db 67 QNVLTAAHCVGN--VVCPEETRVQLREQHLYQD-----HLLPNRIVMHPNYTYPEN 118

QY 137 GADLALLRLASPASLGPAPVWPVCLPRASHRFVHGTAQWATGWDVQOEAADPLPLPWVLQEV 196
Db 119 GADIALLELEDPNVNSAHVQVPTLPPALQTFPTGTCWVTGWDVHSGTLPPLPPFPKQV 178

QY 197 ELRLIGEATCQCLYSQFGFNLTLQLPG-----MLCAGYPEGRDRDTCGDSGGPLVC 249
Db 179 KVPIVENSMCDVQY-----HLGLSTGDGVRIREDMLCAG--NSKSDSCQDGGGGLVC 230

QY 250 EEGGRWFQAGITSFGFCGRRNRPGVFTAVATYEAWIREQVMGSEPGPAFTQP 303
Db 231 RVRGWVLQAGVVSWEHGCAQPNRPGLIYTRVAYYLDWIHQYV-----PKXP 275

Search completed: February 25, 2004, 15:47:56
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 15:36:03 ; Search time 84 Seconds
(without alignments)
1340.952 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKGVLPGLGAVANSDS.....TKSLVLPWLSPLHLLGLWGF 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp archaea:*
2: sp bacteria:*
3: sp fungi:*
4: sp human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archesp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1202	61.5	766	Q8NBY4	Q8NBY4 homo sapien
2	630.5	32.3	339	Q99144	Q99144 mus musculus
3	613	31.4	340	Q8BJV6	Q8BJV6 mus musculus
4	603	30.9	277	Q80WM7	Q80WM7 mus musculus
5	603	30.9	331	Q80X17	Q80X17 mus musculus
6	600	30.7	331	Q8RIA6	Q8RIA6 mus musculus
7	590	30.2	284	Q8NFB6	Q8NFB6 homo sapien
8	590	30.2	327	Q8N171	Q8N171 homo sapien
9	582	29.8	321	Q96RZ8	Q96RZ8 homo sapien
10	570	29.2	328	Q8BJR6	Q8BJR6 mus musculus
11	567	29.0	389	Q9PVX7	Q9PVX7 xenopus lae
12	561	28.7	320	Q7T0X2	Q7T0X2 xenopus lae
13	560.5	28.7	297	Q88781	Q88781 rattus ratt
14	558	28.6	317	Q9DGR3	Q9DGR3 xenopus lae
15	540	27.6	422	Q8WVCL	Q8WVCL homo sapien
16	540	27.6	581	Q9BYE2	Q9BYE2 homo sapien

17	538.5	27.6	307	11	Q7TML0	Q7TML0 mus musculus
18	538.5	27.6	558	4	Q86YM4	Q86YM4 homo sapien
19	537	27.5	328	11	Q80Z40	Q80Z40 rattus norv
20	537	27.5	537	4	Q9BYE1	Q9BYE1 homo sapien
21	530.5	27.2	311	11	Q80XZ3	Q80XZ3 rattus norv
22	530	27.1	855	11	Q9JUI7	Q9JUI7 rattus norv
23	529	27.1	336	11	Q80YD8	Q80YD8 mus musculus
24	521	26.7	471	11	Q8CFE0	Q8CFE0 mus musculus
25	516	26.4	310	11	Q91XC4	Q91XC4 mus musculus
26	510.5	26.1	572	11	Q8BIK6	Q8BIK6 mus musculus
27	510	26.1	310	11	Q9QY29	Q9QY29 mus musculus
28	501	25.7	1059	4	Q7Z411	Q7Z411 homo sapien
29	497.5	25.5	273	6	Q9XSM1	Q9XSM1 ovnis aries
30	495	25.3	453	11	Q812A6	Q812A6 mus musculus
31	493.5	25.3	282	11	Q9D4I3	Q9D4I3 mus musculus
32	493.5	25.3	322	11	Q920S2	Q920S2 mus musculus
33	493	25.2	624	11	Q9DAT3	Q9DAT3 mus musculus
34	491	25.1	855	4	Q7Z410	Q7Z410 homo sapien
35	489	25.0	275	4	Q86TMS	Q86TMS homo sapien
36	489	25.0	276	4	Q86UA5	Q86UA5 homo sapien
37	487	24.9	624	11	Q91Y47	Q91Y47 mus musculus
38	483.5	24.8	806	6	O18783	O18783 macropus eu
39	479	24.5	329	13	O42272	O42272 xenopus lae
40	474.5	24.3	335	11	Q8VIF2	Q8VIF2 mus musculus
41	473.5	24.2	275	4	Q96RZ6	Q96RZ6 homo sapien
42	473	24.2	421	11	Q60491	Q60491 cavia porce
43	472.5	24.2	624	6	Q95ME7	Q95ME7 oryctolagus
44	470.5	24.1	643	6	O97506	O97506 sus scrofa
45	469	24.0	638	11	Q8ROP5	Q8ROP5 mus musculus

ALIGNMENTS

RESULT 1

ID	Q8NBY4	PRELIMINARY;	PRT;	766 AA.
AC	Q8NBY4;			
DT	01-OCT-2002 (TrEMBLrel. 22, Created)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Hypothetical protein FLJ90661.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
KC	TISSUE=Placenta;			
RA	Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,			
RA	Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,			
RA	Yamanoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,			
RA	Masubo Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,			
RA	Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;			
RT	"NEDO human cDNA sequencing project."			
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
DR	EMBL; AK075142; BAC11431.1; -			
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.			
DR	GO; GO:0008233; F:peptidase activity; IEA.			
DR	GO; GO:0004295; F:trypsin activity; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro; IPR009003; Cys Ser trypsin.			
DR	InterPro; IPR001254; Peptidase S1.			
DR	InterPro; IPR001314; Peptidase_S1A.			
DR	Pfam; PF00089; trypsin; 3			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00020; Tryp SPC; 3.			
DR	PROSITE; PS00240; TRYP SIN DOM; 3.			
DR	PROSITE; PS00135; TRYP SIN SER; 1.			
KW	Hypothetical protein; Hydrolase; Protease; Serine protease.			
SQ	SEQUENCE 766 AA; 82564 MW; 3630D550CB06BD55 CRC64;			

Query Match 61.5%; Score 1202; DB 4; Length 766;
 Best Local Similarity 100.0%; Pred. No. 2.1e-88;
 Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 MTNGLEPAEWSVLLGVHSGDPLDGAHTRAVAAIVVPANYSQVELGADLALLRLASPA 149
 DB 1 MTNGLEPAEWSVLLGVHSGDPLDGAHTRAVAAIVVPANYSQVELGADLALLRLASPA 60
 QY 150 SLGPVAVPVCPLPRASHRFVHGACWATGWDVQVQADPLPLPVLQVEVRLIGEATCQCL 209
 DB 61 SLGPVAVPVCPLPRASHRFVHGACWATGWDVQVQADPLPLPVLQVEVRLIGEATCQCL 120
 QY 210 YSQPGPFNLTQLPQMLCAGYFEGARDTCQDSCGGLVCEGGRWFOAGITSPFGGGR 269
 DB 121 YSQPGPFNLTQLPQMLCAGYFEGARDTCQDSCGGLVCEGGRWFOAGITSPFGGGR 180
 QY 270 RNRPGVTAATAYEAWIREQVMSRPGPAFTPOKQTQSD 309
 DB 181 RNRPGVTAATAYEAWIREQVMSRPGPAFTPOKQTQSD 220

RESULT 2

Q99L44 PRELIMINARY; PRT; 339 AA.
 AC Q99L44;
 DT 01-JUN-2001 (TremBLrel. 17, Created)
 DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
 DE Similar to protease, serine, 8 (Prostasin).
 GN PRS8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SvJ, and Swiss; TISSUE=Lung;
 RA Verghese G.M., Caughey G.H.;
 RT "Molecular cloning and characterization of mouse prostasin, a type I
 membrane-associated serine protease of the gamma-tryptase/prostasin
 gene family."
 RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Kitamura K., Takefumi N., Kimio T.;
 RT "mouse serine protease."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; BC003851; AA003851.1; -.
 DR EMBL; AF378086; AAL06320.1; -.
 DR EMBL; AF378085; AAL06319.1; -.
 DR EMBL; AB038244; BAB82496.1; -.
 DR HSSP; P00734; IUVS.
 DR MGD; MGI:1923810; Prss8.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 339 AA; 36216 MW; BC2DE88BC057AF10 CRC64;

Query Match 32.3%; Score 630.5; DB 11; Length 339;
 Best Local Similarity 41.7%; Pred. No. 8.7e-43;
 Matches 145; Conservative 48; Mismatches 138; Indels 17; Gaps 9;
 QY 1 MAQKGVLGPGQLGAVANSDSYSLYGLVPSG-PARG-PPYCGRP-EPSSARIYVGGSSNAQPGT 57
 DB 1 MALRVGLGLQLEAVT---ILLGLLQSGGRADGTREASCGAVTQP--RTTGGGSAKPGQ 55
 QY 58 WPMQVSLHHGGCHT.CGGSLIAPSVLAAAHCFMTNGTLEPAEWSVLLGVHSGDGLDGA 117
 DB 56 WPMQVSLHHGGCHT.CGGSLIAPSVLAAAHCFMTNGTLEPAEWSVLLGVHSGDGLDGA 112
 QY 118 HTRAVAAIVVPANYSQVELGADLALLRLASPASLGPVAVPVCPLPRASHRFVHGACWATG 177
 DB 113 VVHTVAQIITHSSYRESGQGDIALIRLSSPVTFSTYRPIRPICLPAANASFFNGLHCTVTG 172
 QY 178 WGDVQVADPLPLPVLQVEVRLIGEATCQCLYQPGPFNLTQLPQMLCAGYFEGARD 237
 DB 173 WGHVAPSVSLQTPRPLQOQLEVLISRETSCLYNINAVPEEFTIQDMLCAGYVKGKGD 232
 QY 238 TCQDSCGGLVCEGGRWFOAGITSPFGGRRNRPGVFTAVATYEAWEIRQVMSRPGP 297
 DB 233 ACQDSCGGLVCEGGRWFOAGITSPFGGRRNRPGVFTAVATYEAWEIRQVMSRPGP 290
 QY 298 APPTQPKTQSD---CLHQTAFLDAR-ILLRPLSHISVGSTGKSL 341
 DB 291 RVVPTQESQDGLCNHHPVFSSAAAPKLLRPVLPFLGLTGLLSL 338
 RESULT 3
 Q8BJV6 PRELIMINARY; PRT; 340 AA.
 AC Q8BJV6;
 DT 01-MAR-2003 (TremBLrel. 23, Created)
 DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
 DE Prostasin.
 GN PRS8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Eye;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK078696; BAC37362.1; -.
 DR MGD; MGI:1923810; Prss8.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 SQ SEQUENCE 340 AA; 36503 MW; 3C2540E9B3A81C8A CRC64;

Query Match 31.4%; Score 613; DB 11; Length 340;
 Best Local Similarity 41.5%; Pred. No. 2.2e-41;
 Matches 145; Conservative 46; Mismatches 140; Indels 18; Gaps 10;
 QY 1 MAQKGVLGPGQLGAVANSDSYSLYGLVPSG-PARG-PPYCGRP-EPSSARIYVGGSSNAQ-PG 56


```
Db 1 MALRVGLGLQLEAVT---ILLLLGLLQSGIRADGTASCGAVIQP--RITGGGQWQSPG 55
QY 57 TWPWQVSLHHGGGHI CGGSLIAPSWLNAHCPWTNGTLEPAEAWSVLLGVHSDQGLDG 116
Db 56 QMPWQVSIITDGNHVCGLSVNKNWVSAACHCFREHSRE--AYEVKLGDDHQLDSYND 112
QY 117 AHTRAVAALVVPANYSQVELGADLALLRLASPASLGPVAVMPVCLPRASHRFVHGTACWAT 176
Db 113 TVVHTVAQLIITHSSYREEGSQGLALIRLSSPFFSYRIRPICLPANASFPNGLHCTVT 172
QY 177 GNGDVQEAADPLPLPWVQLQVEVELLGEATCCQLYSQPGPNLTLQILPGLMCLACGYPEGR 236
Db 173 GNGHVAPSVSLQPRPLQLELEVLISRETCSCLYN:NAVPEEHTTQQDMCLCAGYVKGK 232
QY 237 DTCQSGGGLVCEEGRWFOAGITSGFCGGRNRPGVETAVATVATVATVATVATVATVATV 296
Db 233 DACQSGGGLSPCMGIMGYLAGVSWGDACGAPNRPVTVTLTSTYASVHHHV--AELQ 290
QY 297 PAFPTQPKTQSD---CLHQTAFEDSAR-ILLRPLSHISVGTGTGKSL 341
Db 291 PRVVPQTESQPDGHLNHHFVSSAAAPKLLRPVLFPLGLTLGLLSL 339
RESULT 4
Q80WM7 PRELIMINARY; PRT; 277 AA.
AC Q80WM7;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Tryptase-6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Testis;
RA Wong G.W., Yasuda S., Li L., Stevens R.L.;
RT "Cloning and characterization of mouse tryptase-6 (mt6).";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY262880; AAF20885.1; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 277 AA; 29887 MW; 3310C067573E5E74 CRC64;
Query Match 30.9%; Score 603; DB 11; Length 277;
Best Local Similarity 46.4%; Pred. No. 1.1e-40;
Matches 122; Conservative 30; Mismatches 89; Indels 22; Gaps 6;
QY 38 CGRPEPSARIVGGSNAQPGTWQVSLHHGGGHI CGGSLIAPSWLNAHCPWTNGTLEP 97
Db 25 CGQPRMSRIVGGRDAQDGEWPMQTSIQHGAHVCGGSLIAPQWVLTAGHCF-----P 77
QY 98 ----AAEWSVLLGVHSDQGLDGHTRAVA--IVVPANYSQVELGADLALLRLASPASL 151
Db 78 RRVMPSEYVLLGALSUD--VRSSHELLVPLVRLPPDYSEDEARGDLALLQLRHFVSL 135
QY 152 GPVAVMPVCLPRASHRFVHGTACWATGVDVQEAADPLPLPWVQLQVEVELLGEATCCQLY- 210
Db 136 STRIQVCLPAGSHPPPGSPCWVTGWSLSPGVPLPKRPGQVVRVPLLLDSRACDLVH 195
```

```
QY 211 ---SOPGPENLTQLIIPGMLCAGYPEGREDTCCQDGGPLVCEEGRWFOAGITSGFGC 267
Db 196 VGANVPOGRI---VLPGLMCLCAGYRGHDKACQDSSGGELTQWESGHWLVGVVSMGKGC 252
QY 268 GRNRPGVETAVATVATVATVATVATVATVATVATVATVATVATVATVATVATVATVAT 290
Db 253 ALPNRPVTVNPKYSPWIAQL 275
RESULT 5
Q80X17 PRELIMINARY; PRT; 331 AA.
AC Q80X17;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Tryptase 5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RA Wong G.W., Yasuda S., Stevens R.L., Li L.;
RT "Cloning and characterization of mouse tryptase-5 (mt5).";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY266139; AAP23216.1; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 331 AA; 35640 MW; B1B36F32EF720377 CRC64;
Query Match 30.9%; Score 603; DB 11; Length 331;
Best Local Similarity 43.9%; Pred. No. 1.4e-40;
Matches 129; Conservative 38; Mismatches 115; Indels 12; Gaps 6;
QY 38 CGRPEPSARIVGGSNAQPGTWQVSLHHGGGHI CGGSLIAPSWLNAHCPWTNGTLEP 97
Db 45 CGRPRTSGRIVSGQDAQLGRWPMQVSVRENGAHVCGGSLIAPEDVLTAAHCFNQGSQ-- 102
QY 98 AAEWSVLLG-VHSDQGLDGHTRAVAALVVPANYSQVE-LGADLALLRLASPASLGPV 155
Db 103 -SIYTVLLGTISYPEDNEPKELRAVAQIKPISYSADEHSSGDIALVQLASISFNIDYM 161
QY 156 WPVCLPRASHRFVHGTACWATGVDVQEAADPLPLPWVQLQVEVELLGEATCCQLYSQ--- 212
Db 162 LPVCLPKPGDPLDPGTMCWVTGWHIGTNQPLPPPTLQELQVPLDAETCTNYQENSI 221
QY 213 PGPFNLTQLIIPGMLCAGYPEGREDTCCQDGGPLVCEEGRWFOAGITSGFGCGRRNR 272
Db 222 PGTEFY---ILRGLMCLCAGFEQKDKACNGDSGGPLVCDINDVMIQAGVYVSGSDCALFKR 278
QY 273 PGVETAVATVATVATVATVATVATVATVATVATVATVATVATVATVATVATVATVAT 326
Db 279 PGYTNVSVIISVI-QNTWNPLMEGRGSPSLSGTPLGLLTSLSAFLFLGP 331
RESULT 6
Q8RIAG PRELIMINARY; PRT; 331 AA.
ID Q8RIAG
AC Q8RIAG;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
```



```
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
SQ SEQUENCE 327 AA; 35106 MW; 60458ED0817AC1CF3 CRC64;
Query Match 30.2%; Score 590; DB 4; Length 327;
Best Local Similarity 46.3%; Pred. No. 1.5e-39;
Matches 119; Conservative 29; Mismatches 99; Indels 10; Gaps 3;
QY 38 CGRPEPARIYCGSNAQPGTWPMQVSLHGGGHICGSLIAPSWLSAAHCFMTN 97
DB 75 CGOPRMRSSIRVGGDRGDGEPWQASIQHGAHVCGLIAPQWLVTAACHFPRAL--- 131
QY 98 AAWSVLLGVHSDGDLGDAHTRAVAIVVPANYSQVELGADLALRLASPLGPAVWP 157
DB 132 PAEYRVRLGALRGSTSPRTLSVYVRRVLLPPDYSEDGARGDLALQLRRPVLSARVQ 191
QY 158 VCLPRASHRFVHGTACWATGWQVQADPLPLPWVLOEVELRLLGATCOCLY---SOP 213
DB 192 VCLPVPGARPPPGTPCRVTGWGSLRPGVLPFWRLQGVVRVPLDSRTCDGLYHVGADVP 251
QY 214 GPENLTQLLPGMLCAGYEGRRDTCQSDGGPLVCEEGRWFFQAGITSFEGCGRRNEP 273
DB 252 QABRI---VLPGSLCAGYQGHKDACQSDGGPLTCLQSGDWLVGVSWGKGCALPNRP 308
QY 274 GVFTAVATYEAWIREQV 290
DB 309 GVYTVATYSPWLOARV 325
RESULT 9
ID Q96RZ8 PRELIMINARY; PRT; 321 AA.
AC Q96RZ8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE HS transmembrane tryptase, gene name TMT, AF175522_1.
GN TRPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096910; PubMed=11157797;
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA Higgs D.R.;
RT "sequence, structure and pathology of the fully annotated terminal 2
RT Mb of the short arm of human chromosome 16.";
RL Hum. Mol. Genet. 10:339-352(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AE006466; AK61269.1; -.
DR HSSP; P00761; 1ANI.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM0020; Tryp_SPC; 1.
DR PROSITE; PSS0240; TRYPSIN_DOM; 1.
DR HYDROLASE; PS00134; TRYPSIN_HIS; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
SQ SEQUENCE 321 AA; 33829 MW; FFF5089EDCA4FC73D CRC64;
Query Match 29.8%; Score 582; DB 4; Length 321;
Best Local Similarity 46.0%; Pred. No. 6.6e-39;
Matches 128; Conservative 35; Mismatches 87; Indels 28; Gaps 11;
QY 36 PYCGRPEPS---ARIVGGSSNAQPGTWPMQVSLHGGGHICGSLIAPSWLSAAHCFMTN 92
DB 24 FCGRGPQVSDAGGRIVGHAAPAGAPWQASLRLRRVHVCGSLSPQWLVTAACHF--S 81
QY 93 GTLEPAEWSVLLGVHSDGDLPLD---GAHTRAVAAIVVPANYS-QVELGADLALRLASP 148
DB 82 GSLN-SSDYQVHL-----GELEITLSPHFSTVRQIILHSSPSGPGTSGDIALVELSVP 134
QY 149 ASLGAFAVPVCLPRASHRFVHGTACWATGWQVQADPLPLPWVLOEVELRLLGATCOC 208
DB 135 VTLSSRIILPVCLPEASDDPCFCIRCWVTGWYTRGEPLPPYSIREVKSVYVDVETCR 194
QY 209 LYSQSGPENLTQLLPGMLCAGYEGRRDTCQSDGGPLVCEEGRWFFQAGITSFEGCGG 268
DB 195 DY--FQPGSIIQ--PDMICA---RPGDACQDGGPLVCQVGNWQAGIVSWGEGCG 247
QY 269 RNRPGVFTAVATYEAWIREQVM---GSEPGPAFTQP 303
DB 248 RNRPGVTRVPAVYVNWIRRHITASGSESG--YPLRP 283
RESULT 10
ID Q8BJR6 PRELIMINARY; PRT; 328 AA.
AC Q8BJR6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Similar to MARAPISIN precursor (Channel-activating protease 2-like
DE protein) (Pancreasin).
GN MPN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354883; PubMed=12466851;
RA The RIKEN Consortium,
RA the FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=12986/SvevTac;
RA Brathwaite M., Waeltz P., Schlessinger D., Nagaraja R.;
RT "Genomic Sequence Analysis in the Mouse T-complex Region.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Bladder;
RX MEDLINE=22437995; PubMed=12441343;
RA Bhagwandin V.J., Hau L.W., Mallen-St Clair J., Wolters P.J.,
RA Caughey G.H.;
RT "Structure and activity of human pancreasin, a novel tryptic serine
RT peptidase expressed primarily by the pancreas.";
RL J. Biol. Chem. 278:3363-3371(2003).
DR EMBL; AK080281; BAC37864.1; -.
DR EMBL; AY162410; AAC17162.1; -.
DR MBL; AF542056; AAC27572.1; -.
DR MGD; MGI:2450123; Mpn.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
```

```
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
SQ SEQUENCE 328 AA; 35789 MW; DCOB20FLAB3EB840 CRC64;

Query Match      29.2%; Score 570; DB 11; Length 328;
Best Local Similarity 41.3%; Pred. No. 6.2e-38;
Matches 131; Conservative 38; Mismatches 126; Indels 22; Gaps 8;

QY 26 LVPSGP--ARGPPYCGRPEPSARIIVGGSNAOQGTWPQVSLHGGCHICGSLIAPSWVL 83
DB 15 LRSSTEGARTLRACHPKFQFNRVGGENALEGEWPQVSIQRNGHFCGSLIAPTWWL 74
QY 84 SAAHCFMTNGTLEPAEASVLGVHSQDGLDGAHTRAVAAIVPAN--YSQVELGADLA 141
DB 75 TAAHCFNSTSDI---SIYQVLLGALKLQOP--GPHALYVFPVKVKSNPQYQGMASADVA 129
QY 142 LLRLASPASLPAVWPVCLPRASHRFVHCTACWATGWDVQADPLPLPWVLOEVELRL 201
DB 130 LVELOQPVTFNIIPLVCLPDSVIFESGMNCWVTGWSGSPSQDLRPNRVLQKLAVPIL 189
QY 202 GEATCQCLYSQPGPNLTIQ--ILPGMLCAGYPEGRDTCQDGGGGLVCEEGGRWFQAGI 260
DB 190 DTPKCNLLYNKDVESDFQKTKIDDMLCAGFAEGKKACKGDSGGLVCLVDSQVQAGV 249
QY 261 TSFGGCGRRNRPVFTAVATYEAWMIREQVMGSE--PGPAFFTPQOK-----TQSD 309
DB 250 ISWGCGCARNRPGVYIRVTSHHKWHQIIPELQFGAGTQQCKQKDSQGOORLAGNSAP 309
QY 310 CLHQTAFL-LDSARILLR 325
DB 310 CLAHMVLALGALLR 326

RESULT 11
Q9PVX7
ID Q9PVX7 PRELIMINARY; PRT; 389 AA.
AC Q9PVX7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Epidermis specific serine protease.
GN XEPSIN.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K.;
RT "The expression control of xepsin by non-axial and planar
RT posteriorizing signals in Xenopus epidermis."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DB EMBL; AB018694; BAA84941.1; -
DB HSSP; P00763; IDPO.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR01254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 389 AA; 42375 MW; B31FB4A2F5D1F6E3 CRC64;

Query Match      29.0%; Score 567; DB 13; Length 389;
Best Local Similarity 40.2%; Pred. No. 1.3e-37;
Matches 109; Conservative 45; Mismatches 97; Indels 20; Gaps 5;

QY 38 CGRPEPSARIIVGGSNAOQGTWPQVSLHGGCHICGSLIAPSWVLSAAHCFMTNGTLEP 97
DB 17 CGVPVSNRIIVGGMDSKRGWFWQVLSYKSDSICGSLITDSWVMTAAHCL---DSLQ- 72
QY 98 AAESVLGVHSQDGLDGAHTRAVAAIVPANYSQVELGADLALLRLASPASLGPVAVP 157
DB 73 VSYITVYLGAQLSAPDNSTVSRGVKSIKHPDFQEGSGDIALIELEKPTFTPIPLP 132
QY 158 VCLPRASHRFVHCTACWATGWDVQADPLPLPWVLOEVELRLGEATCQCLYSQPGPN 217
DB 133 ICLPSQDVQFAAGTMCWVTGNTQEGTFLSPKTIQAEVAIIDSSVCGTMYES----- 187
QY 218 LTLQILP-----GMLCAGYPEGRDTCQDGGGGLVCEEGGRWFQAGITSFGCGRR 270
DB 198 -SLGYIPDFSIQEDMVCAGYKEGIDACQDGGGLVGVNNTVWVWLQGLVSWGVCRAEP 246
QY 271 NRPQVFTAVATYEAWMIREQV---MGSEPGPA 298
DB 247 NRPQVYTKVQYQDWLKNVPLIVFSEEGPS 277

RESULT 12
Q7TOX2
ID Q7TOX2 PRELIMINARY; PRT; 320 AA.
AC Q7TOX2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Embryo;
RC MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative."
RL Dev. Dyn. 225:384-391 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
```


Query Match 28.6%; Score 558; DB 13; Length 317;
Best Local Similarity 39.2%; Pred. No. 5.6e-37;
Matches 116; Conservative 45; Mismatches 119; Indels 16; Gaps 5;

QY 35 PPYGRPPSARIIVGGSSNAQPGTWQVLSLHGGHICGGSLIAPSWLVAHCF---M 90
DB 29 PPLCGSPFSSRIIVGGTITROGAFWQVLSLFNGSHICGGSLISDOWLLTATHCIEHPDL 88
QY 91 TNGTLEPAAEWSVLGVHSQDGLDGAHTRAVAIVPANYQVELGADLALLRIASPAS 150
DB 89 PSGC-----GVLGAYQLVKNPHEMTVKDIIYINSEFNGRTSGDIALLLKSLSPK 141
QY 151 LGPAWVPVCLPRASHRFVHGTACWATGWDVQEADPLPWLQVLELRLGEATCCCLY 210
DB 142 FTEYILPCLPASPTVFSSTGTEMTGQGTGSEVPLQYPATLQKVWPIINRSCENY 201
QY 211 SQPGFNLTLQILPGMLCAGYPEGRRDTCQDGGPLVCEBGRWFOAGITSEFGGCR 269
DB 202 HINSVISETEILIQSDQICAGYQAGQKQCGDGGPLVCKIQGFWYQAGIVSWGERCAA 261
QY 270 RNRPGVFTAVATYEAWIREQ-VMGSEPGAPPTQPKTQSCDLHOTAFIDGARILL 324
DB 262 KNRPGVYTFPAYETWISERSVISFKP---FTSSSPSSSSVLRASAILLGVSLIL 314

RESULT 15

Q8WVC1 PRELIMINARY; PRT; 422 AA.

AC Q8WVC1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC018146; AAH18146.1; -
DR HSSP; P00761; IANL.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 4.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS50068; LDLRA_2; 4.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hypothetical protein; Hydrolase; Protease; Serine protease.
FT NON TER 1
SQ SEQUENCE 422 AA; 46257 MW; 2C99875D1B58B319 CRC64;

Query Match 27.6%; Score 540; DB 4; Length 422;
Best Local Similarity 40.2%; Pred. No. 2.1e-35;
Matches 113; Conservative 39; Mismatches 95; Indels 34; Gaps 7;

QY 32 ARGPPYCGRDEP-----SARIVGGSSNAQPGTWQVLSLH-H-GGGHI 71
DB 148 SKNPECDKEDKSDGSDKDCGLRSFTRQARVVGTTADDEGENPQVLSLHAGQGH 207
QY 72 CGSLIAPSWLVAHCFMTNG---TLEPAAEWSVLGVHSQ-DGPLDGAHTRAVAIVP 128
DB 208 CGASLISPNNLVSAAHCTIDDRGFYSDPTQWTAFLGLHDQSORSAFPGVQERLKRILSH 267
QY 129 ANYSQVELGADLALLRIASPASLGPVAVPVCPLPRASHRFVHGTACWATGWDVQEADPLP 188
DB 268 PFTNDFDYDIALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWYTGHTQYGGTGA 327
QY 189 LPWVLELRLGEATCCCLYQSGPGFNLTLQILPGMLCAGYPEGRRDTCQDGGGGL- 247
DB 328 L--ILQKGEIRVINQTTCCENLLPQ-----QITPRMCMCVGLSGGVDSCQDGGGGLS 377
QY 248 VCEEGGRWFOAGITSEFGGCRNRPGVFTAVATYEAWIRE 288
DB 378 SVEADGRIFQAGVYVWGDGCAQRNKPQVYTRLPFLRFDWIKE 418

Search completed: February 25, 2004, 15:49:33
Job time : 87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 15:45:03 ; Search time 44 Seconds
(without alignments)
418.874 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953
Sequence: 1 MAQKGVLPGLGAVANSDS.....TKSLVLPWLSPHSLGLWGF 357

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pdp.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pdp.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pdp.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pdp.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pdp.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	602	30.8	299	3	US-08-944-483-66
2	590	30.2	284	4	US-09-387-375-7
3	563	28.8	290	4	US-09-386-653A-7
4	557	28.5	316	4	US-09-387-375-9
5	550.5	28.2	319	4	US-09-386-642-12
6	550	28.2	328	4	US-09-386-642-11
7	544	27.9	317	4	US-09-386-629-7
8	544	27.9	317	4	US-09-907-794A-263
9	544	27.9	317	4	US-09-905-125A-263
10	544	27.9	317	4	US-09-902-775A-263
11	540	27.6	855	2	US-09-027-337-2
12	540	27.6	855	4	US-09-644-600-2
13	540	27.6	855	4	US-09-654-600A-2
14	534.5	27.4	314	3	US-09-008-271A-3
15	534.5	27.4	314	4	US-09-907-794A-257
16	534.5	27.4	314	4	US-09-905-125A-257
17	534.5	27.3	314	4	US-09-902-775A-257
18	533.5	27.3	314	4	US-09-023-942A-6
19	531	27.2	315	4	US-09-386-653A-9
20	529	27.1	285	4	US-09-023-942A-26
21	526.5	27.0	312	4	US-09-023-942A-4
22	506.5	25.9	902	4	US-09-644-600-10
23	506.5	25.9	902	4	US-09-644-600A-10
24	502	25.7	327	4	US-09-386-629-8
25	499	25.6	276	2	US-09-016-366A-15
26	499	25.6	276	2	US-08-978-404B-21
27	495	25.3	454	3	US-09-518-046-2

28	491.5	25.2	274	2	US-08-978-404B-5	Sequence 5, Appli
29	491	25.1	235	3	US-08-944-483-65	Sequence 65, Appli
30	490.5	25.1	376	4	US-09-820-002-2	Sequence 2, Appli
31	490.5	25.1	416	2	US-09-000-846-2	Sequence 2, Appli
32	490.5	25.1	417	4	US-09-820-002-4	Sequence 4, Appli
33	487.5	25.0	249	3	US-09-079-970A-5	Sequence 5, Appli
34	487	24.9	306	4	US-09-386-642-53	Sequence 53, Appli
35	485.5	24.9	273	2	US-09-016-366A-19	Sequence 19, Appli
36	485.5	24.9	273	2	US-08-978-404B-14	Sequence 14, Appli
37	484.5	24.8	274	2	US-09-016-366A-21	Sequence 21, Appli
38	484.5	24.8	274	2	US-08-978-404B-16	Sequence 16, Appli
39	482.5	24.7	244	4	US-09-601-318-4	Sequence 4, Appli
40	482.5	24.7	244	4	US-09-601-318-5	Sequence 5, Appli
41	482.5	24.7	244	4	US-09-601-318-6	Sequence 6, Appli
42	482.5	24.7	244	4	US-09-601-318-7	Sequence 7, Appli
43	482.5	24.7	245	3	US-09-079-970A-6	Sequence 6, Appli
44	482.5	24.7	245	4	US-09-601-318-1	Sequence 1, Appli
45	479	24.5	256	2	US-09-027-337-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-944-483-66
; Sequence 66, Application US/08944483
; Patent No. 6232456

; GENERAL INFORMATION:

; APPLICANT: COHEN, MAURICE

; APPLICANT: COLPITTS, TRACEY L.

; APPLICANT: FRIEDMAN, PAULA N.

; APPLICANT: GRANADOS, EDWARD N.

; APPLICANT: KLASS, MICHAEL R.

; APPLICANT: RUSSELL, JOHN C.

; APPLICANT: STEWART, KENT D.

; APPLICANT: STROUPE, STEVEN D.

; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES

; NUMBER OF SEQUENCES: 76

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Abbott Laboratories

; STREET: 100 Abbott Park Road

; CITY: Abbott Park

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/944,483

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Becker, Cheryl L.

; REGISTRATION NUMBER: 35,441

; REFERENCE/DOCKET NUMBER: 6183.US.01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 847/935-1729

; TELEFAX: 847/938-2623

; TELEX:

; INFORMATION FOR SEQ ID NO: 66:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 299 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

Patent No. 6635468

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: Ashkenazi, Avi

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerber, Hanspeter

APPLICANT: Gerritsen, Mary B.

APPLICANT: Goddard, A.

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth, J.

APPLICANT: Kljavin, Ivar J.

APPLICANT: Mather, Jennie P.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann.

APPLICANT: Stewart, Timothy A.

APPLICANT: Tuma, Daniel

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William, I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

TITLE OF INVENTION: Acids Encoding the Same

FILE REFERENCE: 10466-14

CURRENT APPLICATION NUMBER: US 09/907,794A

CURRENT FILING DATE: 2001-07-17

PRIOR APPLICATION NUMBER: PCT/US00/04414

PRIOR FILING DATE: 2000-02-22

PRIOR APPLICATION NUMBER: US 60/143,048

PRIOR FILING DATE: 1999-07-07

PRIOR APPLICATION NUMBER: US 60/145,698

PRIOR FILING DATE: 1999-07-26

PRIOR APPLICATION NUMBER: US 60/146,222

PRIOR FILING DATE: 1999-07-28

PRIOR APPLICATION NUMBER: PCT/US99/20594

PRIOR FILING DATE: 1999-09-08

PRIOR APPLICATION NUMBER: PCT/US99/20944

PRIOR FILING DATE: 1999-09-13

PRIOR APPLICATION NUMBER: PCT/US99/21090

PRIOR FILING DATE: 1999-09-15

PRIOR APPLICATION NUMBER: PCT/US99/21547

; PRIOR APPLICATION NUMBER: PCT/US99/21547

1. THE STATE OF TEXAS, County of EL PASO, do hereby certify that JOSEPH A. GARCIA is the holder of the 100 shares of the COMMON stock of the EL PASO ELECTRIC RAILWAY COMPANY, a corporation organized and existing under the laws of the State of Texas, and that said shares are duly paid for and are not subject to any lien or claim of any person other than the holder thereof.

1. THE STATE OF TEXAS, County of EL PASO, do hereby certify that JOSEPH A. GARCIA is the duly qualified and acting County Clerk of said County.

APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/905/125A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20

Query Match 27.9%; Score 544; DB 4; Length 317;
Best Local Similarity 39.3%; Pred. No. 5.4e-42;
Matches 114; Conservative 41; Mismatches 123; Indels 12; Gaps 4;
7 LGPGQIGAVANSYSILYLVPSPARGPPYCGRPPSARIIVCGSNAPQCTWPNQVSLHH 66
10 LGGGCLGTFSTLLASTATINAAIRPVPPACGKPOOLNRVGGEDSTDSEWPWVSIQK 69
67 GGGHICGSLIATPSVLSSAACHFMTNGTLEPAEWSVLLGVHSDQGLDGAHTRAAV 126
70 NGTHHCAGSLTRSWVITAAHCFKDN--LNKPYLFSVLLGAWQLGNPGRSQKVGVAWE 127
127 VPANYSQVE-IGADLALLRLASPASLGPVAVMPCLPASHRFVHGFTACWATGNGDVQ 185
128 PHPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASITLPHENTHCWISGWSIQD 187
186 PLPLPWVQLVEVRLLEGEATCQCLY----SQGPPNLTQILPGLMCAGYEGGRDTCQ 242
188 PLPHPTQLOKLKVPIDSEVCSHLYNRGAGQGP-----ITEDMLCAGYLEGERDAC 241

243 SGGPLVCEGGRWFQAGITSGFGCCRRNRPVGTAVATVATYEAIRQVVG 292
242 SGGPLMCQVDGAWLLAGIISWEGCAERNRPGVVVISLSAHSRWVEKTVQ 291
RESULT 10
US-09-902-775A-263
Sequence 263, Application US/09902775A
Patent No. 6686451
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/902,775A
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20

```

; PRIOR APPLICATION NUMBER: PCT/US00/002191
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 263
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-902-775A-263

```

[illegible]

RESULT 11
 US-09-027-337-2
 ; Sequence 2, Application US/09027337B
 ; Patent No. 5972616
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; APPLICANT: Tanimoto, Hirotsoshi
 ; TITLE OF INVENTION: TAGD-15: An Extracellular Serine Protease Overexpressed in
 ; TITLE OF INVENTION: Breast and Ovarian Carcinomas
 ; FILE REFERENCE: D6064
 ; CURRENT APPLICATION NUMBER: US/09/027,337B
 ; CURRENT FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 13
 ; SEQ ID NO 2
 ; LENGTH: 855
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Amino acid sequence of TAGD-15 encoded by nucleotides
 ; OTHER INFORMATION: 23 to 2589 of Sequence 1
 ; Patent No. 5972616
 US-09-027-337-2

[illegible]

189	QY	LPWVLFVEVRLLEGATCCOCLYSQGFENITLQILPGMLCAGVPGEGRDTCQDSGGL-	247
761	Db	L--ILQGEIIVQTTCTENLLPQ-----QITPRMVCVGLSGVSCQDSGGLS	810
248	QY	VCEEGRWFGAITSFGCGRRNRPGVFTAVATYEAIRNE	288
811	Db	SVKADGRIFQAGVWSWGDCGAQRKNKPGVYVTLPLFRDWIKE	851

```

RESULT 12
US-09-644-600-2
; Sequence 2, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotschi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: RPT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-2

```

[illegible]

RESULT 13
US-09-654-600A-2
; Sequence 2, Application US/09654600A
; Patent No. 6649741
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/654,600A
; CURRENT FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; 09/027,337


```
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 257
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-907-794A-257

Query Match      27.4%; Score 534.5; DB 4; Length 314;
Best Local Similarity 39.0%; Pred. No. 4e-41;
Matches 110; Conservative 53; Mismatches 100; Indels 19; Gaps 8;

QY 29 SGFAGPPYCGRPEPSARIYVGGNAQPGTWQVSLHHGGHICGGSLIAPSWLSAAHC 88
DB 26 AAPLSGP--CGRRVITSRIVGGDAELGRWPWQCSLRLWDSHVCGVSLLSHRWALTAAHC 83
QY 89 FMTNGTLEPAAEWSVLIG--VHSQDG--PLDGAHTR-AVAAIIVVPANYSQVELG---ADLA 141
DB 84 FETYSDLSDPSGWMVQGLTSMPSFWSLQAYITRYFVSNILSPRY----LGNSPYDIA 139
QY 142 LLRLASPASIGPAWVPVCLPRASHRFVHGTHACWATGWDVQVQADPLPLPWVLOEVELRL 201
DB 140 LVKLSAPVTYTKHIQICLOASTFEFFENRTDCWITGNGYIKEDFALPSHTLOEVQVAIL 199
QY 202 GEATCCQLYSQPGFNLTLQILPMLCAGYEGRRDTCQDSDGGPLVCEGGRWFQAGIT 261
DB 200 NNSMCNHLFLK---YSFRKDIFGDMVACAGNAQGGKDACFGDSGGPLACNKNGLWYQIGV 256
QY 262 SFEGCGRRNRPGVFTAVATYEAWIRE---QVMGSEPGAPF 300
DB 257 SWGVGCGRPNEGVTYINSHHFEMIQLMAQSGMSQPDPSWP 298
```

Search completed: February 25, 2004, 15:51:28
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 15:49:38 ; Search time 641 Seconds
(without alignments)
117.600 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKGVLPGLGAVANS...TKSLVLPWLSPHSLGLGWGF 357

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA: *
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1505	77.1	818	9	US-09-888-615-111
2	690	35.3	343	9	US-09-888-615-111
3	690	35.3	343	14	US-10-176-847-86
4	690	35.3	343	14	US-10-097-340-262
5	690	35.3	343	15	US-10-074-566-125
6	677.5	32.1	339	14	US-10-109-616-2
7	621	31.8	342	15	US-10-051-874-86
8	620	31.7	342	15	US-10-051-874-87
9	590	30.2	280	15	US-10-051-874-26
10	590	30.2	284	13	US-10-041-400A-7
11	590	30.2	284	13	US-10-041-264A-7
12	590	30.2	284	13	US-10-042-091A-7
13	583.5	29.9	285	15	US-10-051-874-89
14	582.5	29.8	346	16	US-10-311-035-8
15	575.5	29.5	255	14	US-10-221-097-36

16	567	29.0	389	15	US-10-074-978A-219	Sequence 219, App
17	563	28.8	290	11	US-09-833-245-1294	Sequence 1294, App
18	563	28.8	290	13	US-10-040-655-7	Sequence 7, Appli
19	563	28.8	290	13	US-10-041-006A-7	Sequence 7, Appli
20	563	28.8	290	14	US-10-028-072-222	Sequence 222, App
21	563	28.8	290	14	US-10-121-049-222	Sequence 222, App
22	563	28.8	290	14	US-10-123-904-222	Sequence 222, App
23	563	28.8	290	14	US-10-140-470-222	Sequence 222, App
24	563	28.8	290	14	US-10-175-746-222	Sequence 222, App
25	563	28.8	290	14	US-10-176-918-222	Sequence 222, App
26	563	28.8	290	14	US-10-176-921-222	Sequence 222, App
27	563	28.8	290	14	US-10-137-865-222	Sequence 222, App
28	563	28.8	290	14	US-10-140-474-222	Sequence 222, App
29	563	28.8	290	14	US-10-142-431-222	Sequence 222, App
30	563	28.8	290	14	US-10-143-114-222	Sequence 222, App
31	563	28.8	290	14	US-10-140-002-222	Sequence 222, App
32	563	28.8	290	14	US-10-142-419-222	Sequence 222, App
33	563	28.8	290	14	US-10-123-262-222	Sequence 222, App
34	563	28.8	290	14	US-10-142-423-222	Sequence 222, App
35	563	28.8	290	14	US-10-121-050-222	Sequence 222, App
36	563	28.8	290	14	US-10-141-755-222	Sequence 222, App
37	563	28.8	290	14	US-10-143-032-222	Sequence 222, App
38	563	28.8	290	14	US-10-123-108-222	Sequence 222, App
39	563	28.8	290	14	US-10-123-236-222	Sequence 222, App
40	563	28.8	290	14	US-10-123-261-222	Sequence 222, App
41	563	28.8	290	14	US-10-140-921-222	Sequence 222, App
42	563	28.8	290	14	US-10-140-928-222	Sequence 222, App
43	563	28.8	290	14	US-10-121-045-222	Sequence 222, App
44	563	28.8	290	14	US-10-123-292-222	Sequence 222, App
45	563	28.8	290	14	US-10-123-903-222	Sequence 222, App

ALIGNMENTS

RESULT 1

US-09-888-615-111
; Sequence 111, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; PRIOR FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 111
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-111

Query Match 77.1%; Score 1505; DB 9; Length 818;
Best Local Similarity 100.0%; Pred. No. 6.1e-122;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	38	CGPEPSARIYVGSNAPQGTWPQVSLHGGGHHICGSLIAPSWVLSAAHCFMTNCTLEP	97
Db	38	CGPEPSARIYVGSNAPQGTWPQVSLHGGGHHICGSLIAPSWVLSAAHCFMTNCTLEP	97
Qy	98	AAEWSVLLGVHSDGPDGCAHTRAVAIVVPANYSOVELGADIALRLASPASIGFAVWP	157
Db	98	AAEWSVLLGVHSDGPDGCAHTRAVAIVVPANYSOVELGADIALRLASPASIGFAVWP	157
Qy	158	VCLPRASHRFVHGTAACWATGWGDVQEADELPLPWLQVELRLLGATCQCCLYQSGPFN	217

; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 262
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-262

Query Match 35.3%; Score 690; DB 14; Length 343;
Best Local Similarity 43.6%; Pred. No. 1.1e-51;
Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;
QY 1 MAQKGVLPGLGAVANSDSYSLY-GLVPSGP-----ARGPPYCGRPESARIVGSGNAQP 55
DB 1 MAQKGVLPGLGAVANSDSYSLY-GLVPSGP-----ARGPPYCGRPESARIVGSGNAQP 55
QY 56 GTPWQVSLHGGGHIHCGSLIAPSWLSAAHCFWNTGLEPAAEWSVLLGVHSDGGLD 115
DB 54 GTPWQVSLHGGGHIHCGSLIAPSWLSAAHCFWNTGLEPAAEWSVLLGVHSDGGLD 115
QY 116 GAHTRAVAAIVVPANYSQVELGADLALLRLASGLPAVMPVCLPRASHRFVHGTACWA 175
DB 111 DAKVSTLKDIIHPHSYLGQSGDIALQLSRITFSYRIRPICLPAAANASFPNGLHCTV 170
QY 176 TQMGVDQVADPLPLPWLQVELELRLGEATCCQLYSQPGFNLTLQILPGMLCAGYPEG 235
DB 171 TQMGHVASVSLTTPKPLQQLVPLISRETCLNLYNIDAKPEEPHFVQEDMVCAGYVEGG 230
QY 236 RDTCCGDSGGLPVCBEGGRWFQAGITSFGCGRRNRPQVFTAVATYEAWAREQVWGSEP 295
DB 231 KDACQDSGGLPVCBEGGRWFQAGITSFGCGRRNRPQVFTAVATYEAWAREQVWGSEP 295
QY 296 GFAPPTQPKTQSD---CLHQTAFDLS-ARILLRPLSHISVGSTGTGKSLVLPWLSPH 349
DB 289 QPRVVPQTQESQPSDNLGSHLAFSSAPAGQLLRPLFLPLGLAIG---LLSPWLSEH 343

RESULT 5
US-10-074-566-125
; Sequence 125, Application US/10074566
; Publication No. US20030207348A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same
; FILE REFERENCE: 15966-556 CIP1
; CURRENT APPLICATION NUMBER: US/10/074,566
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: 09/619,252
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/144,722
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/167,785
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 60/276,994
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/280,898
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/332,241
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/288,062
; PRIOR FILING DATE: 2001-05-02

; PRIOR APPLICATION NUMBER: 60/291,766
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/314,007
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 125
; LENGTH: 343
; TYPE: PRT
; ORGANISM: human
US-10-074-566-125

Query Match 35.3%; Score 690; DB 15; Length 343;
Best Local Similarity 43.6%; Pred. No. 1.1e-51;
Matches 156; Conservative 49; Mismatches 129; Indels 24; Gaps 10;
QY 1 MAQKGVLPGLGAVANSDSYSLY-GLVPSGP-----ARGPPYCGRPESARIVGSGNAQP 55
DB 1 MAQKGVLPGLGAVANSDSYSLY-GLVPSGP-----ARGPPYCGRPESARIVGSGNAQP 55
QY 56 GTPWQVSLHGGGHIHCGSLIAPSWLSAAHCFWNTGLEPAAEWSVLLGVHSDGGLD 115
DB 54 GTPWQVSLHGGGHIHCGSLIAPSWLSAAHCFWNTGLEPAAEWSVLLGVHSDGGLD 115
QY 116 GAHTRAVAAIVVPANYSQVELGADLALLRLASGLPAVMPVCLPRASHRFVHGTACWA 175
DB 111 DAKVSTLKDIIHPHSYLGQSGDIALQLSRITFSYRIRPICLPAAANASFPNGLHCTV 170
QY 176 TQMGVDQVADPLPLPWLQVELELRLGEATCCQLYSQPGFNLTLQILPGMLCAGYPEG 235
DB 171 TQMGHVASVSLTTPKPLQQLVPLISRETCLNLYNIDAKPEEPHFVQEDMVCAGYVEGG 230
QY 236 RDTCCGDSGGLPVCBEGGRWFQAGITSFGCGRRNRPQVFTAVATYEAWAREQVWGSEP 295
DB 231 KDACQDSGGLPVCBEGGRWFQAGITSFGCGRRNRPQVFTAVATYEAWAREQVWGSEP 295
QY 296 GFAPPTQPKTQSD---CLHQTAFDLS-ARILLRPLSHISVGSTGTGKSLVLPWLSPH 349
DB 289 QPRVVPQTQESQPSDNLGSHLAFSSAPAGQLLRPLFLPLGLAIG---LLSPWLSEH 343

RESULT 6
US-10-109-616-2
; Sequence 2, Application US/10109616
; Publication No. US2003016748A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CHANNEL
; TITLE OF INVENTION: ACTIVATING PROTEASE 1 (CAPI) GENE DISRUPTIONS
; FILE REFERENCE: R-490
; CURRENT APPLICATION NUMBER: US/10/109,616
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/280,509
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/311,055
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-109-616-2

Query Match 32.1%; Score 627.5; DB 14; Length 339;
Best Local Similarity 41.4%; Pred. No. 2.9e-46;
Matches 144; Conservative 48; Mismatches 139; Indels 17; Gaps 9;
QY 1 MAQKGVLPGLGAVANSDSYSLYGLVPSG-PARG-PPYCGRP-EPASARIVGSGNAQPGT 57
DB 1 MAPRVGLGLGLEAVT---ILLLLGLLQSGTRADGTASCGAVTQP--RITGGGSAKPGQ 55
QY 58 WPWQVSLHGGGHIHCGSLIAPSWLSAAHCFWNTGLEPAAEWSVLLGVHSDGGLDGA 117

Db 56 WPMQVSTYDGNHVGCGSLVSNKVVWAAHCFPREHRE---AYEVKGAHQLDYSYNDT 112
Qy 118 HTRAVAAIVVPANYSQVELGADLALLRLASPASLGPAWVPCVLPRAHFRFVHGTACWATG 177
Db 113 VVHTVAQIIHSSYREEGSQDIAFIRLSSDVTFSRVPICLPAAANASFPNGLHCTVTG 172
Qy 178 WGDVQADPLPLPWVLOVEVRLIGRATCOCLYSQPGFNLTLQILPGMLCAGYPEGRRD 237
Db 173 WGHVAPSVSLQTPRPLQOQLEVPISRETCSCLYNINAVPEEPHTIQDMLCAGYVKGKD 232
Qy 238 TCQDGGGGLVCEEGGRWFQAGITSFGCGRRNRPGVFTAVATYEAWIREQVMSSEPGP 297
Db 233 ACQDGGGGLVCEEGGRWFQAGITSFGCGRRNRPGVFTAVATYEAWIREQVMSSEPGP 297
Qy 298 AFPTQPKTQSD---CLHQTAF-LDSARILLRPLSHISVGVSTGYKSLVLPWL 346
Db 291 RVVPQTQESQDPDGLCNHHPVFSSAAAPKLLRPVLFPLGLTGLLSL 338

RESULT 7

US-10-051-874-86
; Sequence 86, Application US/10051874
; Publication No. US20040005557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spyttek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenooy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Theeruv, Velizar T
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glennda
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Herrman, John L
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409

; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 86
; LENGTH: 342
; TYPE: PR1
; ORGANISM: Rattus norvegicus
; US-10-051-874-86

Query Match 31.8%; Score 621; DB 15; Length 342;
Best Local Similarity 40.8%; Pred. No. 1.1e-45;
Matches 144; Conservative 50; Mismatches 139; Indels 20; Gaps 9;

Qy 1 MAQGVILGPGQLGAVANSYSYLYGLVPS--GPARGPPYCGRP-EPASRIYVGGSSNAOQPGT 57
Db 1 MALRVGLGLGQLEALF---VLLGLLQSRIGADGTRASCAGAVIQP--RITGGGSAKPGQ 55
Qy 58 WPMQVSLHHGGHICGGSLIAPSWVLAAHCFMTNGTLEPAEWSVLGVHVSQDGPLDGA 117
Db 56 WPMQVSIITGVHVCGLGSLVSNQWVVAHCFPREHSE---EYEVKGAHQLDYSFSDI 112
Qy 118 HTRAVAAIVVPANYSQVELGADLALLRLASPASLGPAWVPCVLPRAHFRFVHGTACWATG 177
Db 113 VVHTVAQIIHSSYREEGSQDIALIRLSSPVTFSRVPICLPAAANASFPNGLHCTVTG 172
Qy 178 WGDVQADPLPLPWVLOVEVRLIGRATCOCLYSQPGFNLTLQILPGMLCAGYPEGRRD 237
Db 173 WGHVAPSVSLQTPRPLQOQLEVPISRETCSCLYNINAVPEEPHTIQDMLCAGYVKGKD 232
Qy 238 TCQDGGGGLVCEEGGRWFQAGITSFGCGRRNRPGVFTAVATYEAWIREQVMSSEPGP 297
Db 233 ACQDGGGGLVCEEGGRWFQAGITSFGCGRRNRPGVFTAVATYEAWIREQVMSSEPGP 297
Qy 298 AFPTQPKTQSD---CLHQTAF-LDSARILLRPLSHISVGVSTGYKSLVLPWL 346
Db 291 RVVPQTQESQDPDGLCNHHPVFSSAAAPKLLRPVLFPLGLTGLLSL---WL 340

RESULT 8

US-10-051-874-87
; Sequence 87, Application US/10051874
; Publication No. US20040005557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spyttek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenooy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Theeruv, Velizar T
; APPLICANT: Zerhusen, Bryan D

APPLICANT: Millet, Isabelle
APPLICANT: Miller, Charles E
APPLICANT: Lepley, Denise M
APPLICANT: Smithson, Glennda
APPLICANT: Baumgartner, Jason C
APPLICANT: Herrman, John L
APPLICANT: Peyman, John A
APPLICANT: Gorman, Linda
APPLICANT: Mezes, Peter D
APPLICANT: Kekuda, Ramesh
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Gerlach, Valerie
APPLICANT: Grosse, William M
APPLICANT: Liu, Xiaohong
APPLICANT: Ellerman, Karen
APPLICANT: Rothenberg, Mark
APPLICANT: Stone, David J
APPLICANT: Burgess, Catherine E
TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
FILE REFERENCE: 21402-245
CURRENT APPLICATION NUMBER: US/10/051,874
CURRENT FILING DATE: 2002-09-25
PRIOR APPLICATION NUMBER: 60/268,595
PRIOR FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: 60/325,306
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/262,587
PRIOR FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: 60/272,409
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/262,454
PRIOR FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: 60/276,777
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/291,672
PRIOR FILING DATE: 2001-05-17
PRIOR APPLICATION NUMBER: 60/330,336
PRIOR FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: 60/265,530
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/261,376
NUMBER OF SEQ ID NOS: 269
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 87
LENGTH: 342
TYPE: PRT
ORGANISM: Rattus norvegicus
US-10-051-874-87
Query Match 31.7%; Score 620; DB 15; Length 342;
Best Local Similarity 40.8%; Pred. No. 1.3e-45;
Matches 144; Conservative 50; Mismatches 139; Indels 20; Gaps 9;
Qy 1 MAQGVLPGLGAVANSDSYSLGLVPS--GPARGPPYCRP--EPSARIYGGSSNAOPGT 57
Db 1 MALRVGLGLGQLHALF---ILLGLIGLSRIGADGTASCGAVIQP--RTGGGSAKPGQ 55
Qy 58 WPMQVSLHGGGHICGGSLIAPSWLSAAHCFMTNGTLEPAAEWSVLGVHSDQGPLDGA 117
Db 56 WPMQVSIYGVHVGCGSLVSNQWVSAHCFPREHSKE---EYEVKLGHAHQLDSEFNDI 112
Qy 118 HTRAAVAIWPANYQSVGLGADLALLRLASPASLGPAPWPCVLPASHRFVHGTACWATG 177
Db 113 VVHTVAQIISHSYREEGSQGDIALIRLSSPVTFRSYRIRPCLPANASFPNGLHCTVTG 172
Qy 178 WCDVQEAAPLPLPWVQLQVELLGEATCCCLYSQPGFNLTLQLPCMLCAGYEGRRD 237
Db 173 WGHVAPSLSLQTPRFLQQLQVLEPLISRETCSCLDYNNAVPEPHITQQDMLCAGYVKGKD 232
Qy 238 TCQGDGGGLVCEEGGRWFQAGITGFGCGRRNRPGVFTAVATYEAWIREQVWMSFPGP 297

Db 233 ACQGDGSGPLSCPTIDGLWYLAGIVSGDACGAPNRPQVYTTSTYASWIIHHV--AELOP 290
Qy 298 APTPOKOTQSD---CLHOTAF-LDSARILRLPLSHISGVSTGTKSLVLPWL 346
Db 291 RAVTQTESQPDGHLCHNHPVFNLAQAQKLSRPILFLPLSLTLGLFSL---WL 340
RESULT 9
US-10-051-874-26
Sequence 26, Application US/10051874
Publication No. US20040005557A1
GENERAL INFORMATION:
APPLICANT: Padigaru, Muralidhara
APPLICANT: Alsobrook II, John P
APPLICANT: Colman, Steven D
APPLICANT: Spytek, Kimberly A
APPLICANT: Boldog, Ferenc
APPLICANT: Vernet, Corine AM
APPLICANT: Li, Li
APPLICANT: Shenoy, Suresh G
APPLICANT: Casman, Stacie J
APPLICANT: Guo, Xiaojia Sasha
APPLICANT: Edinger, Shlomit R
APPLICANT: MacDougall, John R
APPLICANT: Malyankar, Uriel M
APPLICANT: Patturajan, Meera
APPLICANT: Shinkets, Richard A
APPLICANT: Pena, Carol EA
APPLICANT: Tchernev, Velizar T
APPLICANT: Zerhusen, Bryan D
APPLICANT: Millet, Isabelle
APPLICANT: Miller, Charles E
APPLICANT: Lepley, Denise M
APPLICANT: Smithson, Glennda
APPLICANT: Baumgartner, Jason C
APPLICANT: Herrman, John L
APPLICANT: Peyman, John A
APPLICANT: Gorman, Linda
APPLICANT: Mezes, Peter D
APPLICANT: Kekuda, Ramesh
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Gerlach, Valerie
APPLICANT: Grosse, William M
APPLICANT: Liu, Xiaohong
APPLICANT: Ellerman, Karen
APPLICANT: Rothenberg, Mark
APPLICANT: Stone, David J
APPLICANT: Burgess, Catherine E
TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
FILE REFERENCE: 21402-245
CURRENT APPLICATION NUMBER: US/10/051,874
CURRENT FILING DATE: 2002-09-25
PRIOR APPLICATION NUMBER: 60/268,595
PRIOR FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: 60/325,306
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/262,587
PRIOR FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: 60/272,409
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/262,454
PRIOR FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: 60/276,777
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/291,672
PRIOR FILING DATE: 2001-05-17
PRIOR APPLICATION NUMBER: 60/330,336
PRIOR FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: 60/265,530
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/261,376
PRIOR FILING DATE: 2001-01-16

```

; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-051-874-26

Query Match      30.2%; Score 590; DB 15; Length 280;
Best Local Similarity 46.3%; Pred. No. 4.2e-43;
Matches 119; Conservative 29; Mismatches 99; Indels 10; Gaps 3;

QY 38 CGPEPSARIYGGNAQPGTWPQVSLHGGGHCIGGSLIAPSWVLSAAHCFMTNGTLEP 97
Db 28 CGQPRMSRIVGGDRDGEWPFQASIQHPGAHVCGGSLIAPQWVLTAAHCFPRAL--- 84
QY 98 AAESVLLGVHSDQGLDGAHTRAVAAIVVPANYSQVELGADLALLRLASPASIGPAVWP 157
Db 85 PAEYRVLGAIRLGSTSPRTLSVFPVRVLLPPDYSEDGARGDLALLQLRRPVPISARVQP 144
QY 158 VCLPRASHRFVHGTAATGWDVQADPLPLPWVQLVEVLRLIGEATCCQCLY-----SQP 213
Db 145 VCLPVGARPPPGTPCRVTGWSLRPGVLPPEWRPLOGVRVPLDLSRTCDGLYHVGADVP 204
QY 214 GPFNLTLQILPGLMCAGYPEGRRDTCQGDSCGPLVCEEGGWFOAGITSFGFGCGRRNRP 273
Db 205 QAERI---VLPGLSCAGYPQGHKDACQDSCGGLTCLQSGSWVLVGVVSWGKGCALPNRP 261
QY 274 GVFTAVATYEAWIREQV 290
Db 262 GYVTSVATYSPWIOARV 278

RESULT 11
US-10-041-264A-7
; Sequence 7, Application US/10041264A
; Publication No. US20020142446A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Oi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,264A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-264A-7

Query Match      30.2%; Score 590; DB 13; Length 284;
Best Local Similarity 46.3%; Pred. No. 4.3e-43;
Matches 119; Conservative 29; Mismatches 99; Indels 10; Gaps 3;

QY 38 CGPEPSARIYGGNAQPGTWPQVSLHGGGHCIGGSLIAPSWVLSAAHCFMTNGTLEP 97
Db 28 CGQPRMSRIVGGDRDGEWPFQASIQHPGAHVCGGSLIAPQWVLTAAHCFPRAL--- 84
QY 98 AAESVLLGVHSDQGLDGAHTRAVAAIVVPANYSQVELGADLALLRLASPASIGPAVWP 157
Db 85 PAEYRVLGAIRLGSTSPRTLSVFPVRVLLPPDYSEDGARGDLALLQLRRPVPISARVQP 144
QY 158 VCLPRASHRFVHGTAATGWDVQADPLPLPWVQLVEVLRLIGEATCCQCLY-----SQP 213
Db 145 VCLPVGARPPPGTPCRVTGWSLRPGVLPPEWRPLOGVRVPLDLSRTCDGLYHVGADVP 204
QY 214 GPFNLTLQILPGLMCAGYPEGRRDTCQGDSCGPLVCEEGGWFOAGITSFGFGCGRRNRP 273
Db 205 QAERI---VLPGLSCAGYPQGHKDACQDSCGGLTCLQSGSWVLVGVVSWGKGCALPNRP 261
QY 274 GVFTAVATYEAWIREQV 290
Db 262 GYVTSVATYSPWIOARV 278

RESULT 12
US-10-042-091A-7
; Sequence 7, Application US/10042091A
; Publication No. US20020142447A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Oi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/042,091A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-400A-7

Query Match      30.2%; Score 590; DB 13; Length 284;
Best Local Similarity 46.3%; Pred. No. 4.3e-43;
Matches 119; Conservative 29; Mismatches 99; Indels 10; Gaps 3;

QY 38 CGPEPSARIYGGNAQPGTWPQVSLHGGGHCIGGSLIAPSWVLSAAHCFMTNGTLEP 97
Db 28 CGQPRMSRIVGGDRDGEWPFQASIQHPGAHVCGGSLIAPQWVLTAAHCFPRAL--- 84
QY 98 AAESVLLGVHSDQGLDGAHTRAVAAIVVPANYSQVELGADLALLRLASPASIGPAVWP 157
Db 85 PAEYRVLGAIRLGSTSPRTLSVFPVRVLLPPDYSEDGARGDLALLQLRRPVPISARVQP 144
QY 158 VCLPRASHRFVHGTAATGWDVQADPLPLPWVQLVEVLRLIGEATCCQCLY-----SQP 213
Db 145 VCLPVGARPPPGTPCRVTGWSLRPGVLPPEWRPLOGVRVPLDLSRTCDGLYHVGADVP 204
QY 214 GPFNLTLQILPGLMCAGYPEGRRDTCQGDSCGPLVCEEGGWFOAGITSFGFGCGRRNRP 273
Db 205 QAERI---VLPGLSCAGYPQGHKDACQDSCGGLTCLQSGSWVLVGVVSWGKGCALPNRP 261
QY 274 GVFTAVATYEAWIREQV 290
Db 262 GYVTSVATYSPWIOARV 278

RESULT 10
US-10-041-400A-7
; Sequence 7, Application US/10041400A
; Publication No. US20020110895A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Oi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,400A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-400A-7

Query Match      30.2%; Score 590; DB 13; Length 284;
Best Local Similarity 46.3%; Pred. No. 4.3e-43;
Matches 119; Conservative 29; Mismatches 99; Indels 10; Gaps 3;

QY 38 CGPEPSARIYGGNAQPGTWPQVSLHGGGHCIGGSLIAPSWVLSAAHCFMTNGTLEP 97
Db 28 CGQPRMSRIVGGDRDGEWPFQASIQHPGAHVCGGSLIAPQWVLTAAHCFPRAL--- 84
QY 98 AAESVLLGVHSDQGLDGAHTRAVAAIVVPANYSQVELGADLALLRLASPASIGPAVWP 157
Db 85 PAEYRVLGAIRLGSTSPRTLSVFPVRVLLPPDYSEDGARGDLALLQLRRPVPISARVQP 144
QY 158 VCLPRASHRFVHGTAATGWDVQADPLPLPWVQLVEVLRLIGEATCCQCLY-----SQP 213
Db 145 VCLPVGARPPPGTPCRVTGWSLRPGVLPPEWRPLOGVRVPLDLSRTCDGLYHVGADVP 204
QY 214 GPFNLTLQILPGLMCAGYPEGRRDTCQGDSCGPLVCEEGGWFOAGITSFGFGCGRRNRP 273
Db 205 QAERI---VLPGLSCAGYPQGHKDACQDSCGGLTCLQSGSWVLVGVVSWGKGCALPNRP 261
QY 274 GVFTAVATYEAWIREQV 290
Db 262 GYVTSVATYSPWIOARV 278
```

```

; SEQ ID NO 7
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-042-091A-7

Query Match
Best Local Similarity 30.2%; Score 590; DB 13; Length 284;
Matches 119; Conservative 29; Mismatches 99; Indels 10; Gaps 3;

QY 38 CGRPEARSIVGSSNAQGTWQVSLHGGGHCIGSLIAPSWLSAAHCFMTNGTLEP 97
Db 28 CGQPMSSRTVGRDGDGEWPMQASIQHPGARVCGSLIAPQWVLTAAHCFPRAL--- 84

QY 98 AAWSVLLGVHSGDGLDGAHTRAVAAIVVPANYSOVELGADLALLRLASPSLIGPAVWP 157
Db 85 PAEYRVRLGALRLGSTSPRTLSVPRVRLPPDYSEDGARGDLALQLRRFVLSRVQP 144

QY 158 VCLPRASHRFVHGTACWATGWDVQVADPLPLPWLQVEVRLRLGEATCCCLY----SQP 213
Db 145 VCLPVPGARPPGTGCRVTCWGLRPGVPLPEWRPLQGVVRVPLDSDRTCDGLYHVGADVP 204

QY 214 GPNLTLQLPGLMCLAGYPPGRDTCQDSDGGLVCEGGRWFQAGITSGFGGGRNRP 273
Db 205 QAERI---VLPGLSLCAGYPPQGHKDACQDSDGGLTCLQSGSVLWVGVSWGKCALPNRP 261

QY 274 GVFTAVATYEAWIREQV 290
Db 262 GVYISVATYSPWQARV 278

```

```

RESULT 13
US-10-051-874-89
; Sequence 89, Application US/10051874
; Publication No. US2004000557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytsek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaohua Saeha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glenda
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Herrman, John L
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

```

```

; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 89
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-051-874-89

```

```

Query Match 29.9%; Score 583.5; DB 15; Length 285;
Best Local Similarity 43.3%; Pred. No. 1.6e-42;
Matches 127; Conservative 39; Mismatches 116; Indels 11; Gaps 6;

QY 1 MAOKGVLPQGLGAVANSYSYLYLVPSC-PARG-PPYCGRP-EPSARVGGSNAPQGT 57
Db 1 MALRVGLGLQLEAVT---ILLLLGLLQSGIRADGTASCNAVTP--RITGGGSAKPGQ 55

QY 58 WPMQVSLHGGGHCIGSLIAPSWLSAAHCFMTNGTLEPAEWSVLLGVHSGDGLDGA 117
Db 56 WPMQVSLYDGNHVCVGGSLVSNKVVSAHCFPREHRE---AYEVKLGAHQLDSYNDT 112

QY 118 HTRAVAAIVVPANYSOVELGADLALLRLASPSLIGPAVWPVCLPRASHRFVHGTACWATG 177
Db 113 VVHTVAQIIITHSSYREBSQGDIALIRLSFVTFSYRIPICLPAANASFPNGLHCTVTG 172

QY 178 WGVQVQVADPLPLPWLQVEVRLRLGEATCCCLYSQCPGFNLTLQLPGLMCLAGYPPGRD 237
Db 173 WGVVAPSVSLQTPRPLQQLLEVPLISRETCLVNNINAVPEEPHTIQODMLCAGYVKGKD 232

QY 238 TCQDSDGGLVCEBGGRWFOAGITSGFGGGRNRPQVFTAVATYEAWIREQV 290
Db 233 ACQDSDGGLVCEBGGRWFOAGITSGFGGGRNRPQVFTAVATYEAWIREQV 285

```

```

RESULT 14
US-10-311-035-8
; Sequence 8, Application US/10311035
; Publication No. US2004002243A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: ELLIOTT, Vicki
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: DELEGEANE, Angelo M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: NGUYEN, Darniel B.
; APPLICANT: LEE, Ernestine A.

```



```

; APPLICANT: HAFALIA, April
; APPLICANT: KHAN, Farrah A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: YAO, Monique G.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: TANG, Y. Tom
; APPLICANT: WALSH, Roderick T.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LU, Yan
; APPLICANT: RAMKUMAR, Jayalaximi
; APPLICANT: XU, Yuming
; APPLICANT: REDDY, Roopa
; APPLICANT: DAS, Depodriya
; APPLICANT: KEARNEY, Liam
; APPLICANT: KALLICK, Deborah A.
; TITLE OF INVENTION: Proteases
; FILE REFERENCE: PI-0123 PCT
; CURRENT APPLICATION NUMBER: US/10/311,035
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 60/212,336; 60/213,955; 60/215,396; 60/216,821; 60/218,946
; PRIOR FILING DATE: 2000-06-16; 2000-06-22; 2000-06-29; 2000-07-07; 2000-07-14
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US2004002343A1 2256251CD1
US-10-311-035-8

```

```

Query Match      29.8%; Score 582.5; DB 16; Length 346;
Best Local Similarity 43.8%; Pred. No. 2.4e-42;
Matches 133; Conservative 38; Mismatches 104; Indels 29; Gaps 12;

QY 10 GQLGAVANDSVLSYGLVSPGARGPPYCGREPS---ARIVGSSNAQPGTWPQVSLHH 66
DB 24 GQ-GRLLHGSAGVFLGSPPTSSFDLGCGRPGQVSDAGRIVGGHAAPAGAMPWQASLRL 82
QY 67 GGCHICGSLIAPSWLSAAHCFMTNGTLEPAAEWSVLGLVHSDQGLD---GAHTRAVA 123
DB 83 RRVHVCGLSLLSQWLTAACHF--SSSLN-SSDYQVHL-----GELETLSPHFSTVR 133
QY 124 AIYVVPANYS--QVELGADLALLRLASPASLGPAWVPVCLPRASHRFVHGTAACWATGWDVQ 182
DB 134 QIILHSSPSGPGTSGDIALVELSVPTLFSRLPVCLPEASDDFCFGIRCWVTGWYTR 193
QY 183 EADPLPLPWVLOVELELLGEATCCLYSOPGPNLTLOILPGMLCAGYPEGRRDTCQGD 242
DB 194 EGEPLPPYSLREVKVSVDVDTETCRRDY--PGPGGSLQ--PMLCA---RGPGDACQDD 246
QY 243 SGGPLVCEGGRWFOAGITSFQFCGGRNRNRPVFTAVATYEAWIREQVM---GSEPGPAF 299
DB 247 SGGPLVCQVNGAWTQAGIVSWGEGCGRPNRPVGVTRYVPAYVNWIRRHITASGGSESG--Y 304
QY 300 PTQP 303
DB 305 PRLP 308

```

```

RESULT 15
US-10-221-097-36
; Sequence 36, Application US/10221097
; Publication No. US20030144476A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdoch, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoqing
; TITLE OF INVENTION: NOVEL COMPOUNDS

```

```

; FILE REFERENCE: GP50016
; CURRENT APPLICATION NUMBER: US/10/221,097
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: PCT/US01/07143
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/187,107
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: 60/236,874
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/188,916
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/237,846
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-221-097-36

Query Match      29.5%; Score 575.5; DB 14; Length 255;
Best Local Similarity 43.7%; Pred. No. 6.8e-42;
Matches 128; Conservative 26; Mismatches 88; Indels 51; Gaps 9;

QY 47 IVGSSNAQPGTWPQVSLHGGGHCIGGSLIAPSWLSAAHCFMTNGTLEPAAEWSVLG 106
DB 1 IVGSSNAQPGTWPQVSLHGGGHCIGGSLIAPSWLSAAHCFMTG-----46
QY 107 VHSQDGLDGAHTRAVAIVVPAN-----YSQVELGADLALLRLASPASLGPAWVPVCLP 161
DB 47 --RQVRCPETRRTRS---ALFTRKRREAYNHYSQSDALLQLAHPT---HTPLCLP 96
QY 162 RASHEFVHGTAACWATGWG-DVQEADPL-PLPWLOEVELRLIGEATCCLYSQPGPFNLT 219
DB 97 QPAHFPPFGASCWATGWDQDTSAPSLSPAGTLENLRLISRTNCIYNQLHORHLS 156
QY 220 LQILFGLMCAGYPEGRRDTCQSDSGPLVCBEGGWFQAGITSFQFCGGRNRNRPVFTAV 279
DB 157 NPAPFGMLCGGPGQPGVQGCGLFGAPLVHEVRGTWFLAGLHSGFDACQGPAPPAVFTAL 216
QY 280 ATYE-----AWIREQVMGSEPGPAFPQKTCSDCLHQTAFDLSARILLRPLS 328
DB 217 PAMRTGSAVWTRQVYFABEP-----EFEAEPGSCLAN-----IRPFS 253

```

Search completed: February 25, 2004, 16:08:53
Job time : 642 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 15:51:34 ; Search time 3779 Seconds
(without alignments)
4094.592 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKVLGPGQLGAVANSDS.....TKSLVLPWLSPHSLIGLWGF 357

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlbp
-Q=/cgn2_1/USPTO.spool_p/US10037417/runat_24022004_101038_17447/app_query.fasta_1.519
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptco -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417 @CGN_1_1_4958 @runat_24022004_101038_17447 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*

- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	DB ID	Description
1	1953	100.0	1102	6	AX512287 Sequence
2	1953	100.0	1102	6	AX512289 Sequence
3	1505	77.1	2457	6	AX360096 Sequence
4	1505	77.1	2681	6	AX480935 Sequence
5	1492	76.4	2810	6	BD127529 Primer fo
6	1492	76.4	2810	9	AK075142 Homo sapi
7	970	49.7	537	6	AX342934 Sequence
8	861	44.1	670	6	BD125219 Primer fo
9	861	44.1	670	6	BD126485 Primer fo
10	799	40.9	456	6	AX342936 Sequence
11	716.5	36.7	127769	9	AC009088 Homo sapi
12	697	35.7	1796	6	AX098193 Sequence
13	697	35.7	1800	9	BC001462 Homo sapi
14	697	35.7	1834	6	AX335777 Sequence
15	697	35.7	1834	6	AX336076 Sequence
16	697	35.7	1834	6	AX336340 Sequence
17	697	35.7	1834	6	AX409567 Sequence
18	697	35.7	1834	6	AX474697 Sequence
19	697	35.7	1834	6	HUMPROS
20	697	35.7	1835	6	AX098215 Sequence
21	658	33.7	232119	2	AC106629 Rattus no
22	652	33.4	1726	6	AX675579 Sequence
23	630.5	32.3	1727	10	AY335911 Mus muscu
24	630.5	32.3	1743	10	BC003851 Mus muscu
25	630.5	32.3	1856	10	AB038244 Mus muscu
26	630.5	32.3	1870	10	AF378085 Mus muscu
27	627.5	32.1	1797	10	AF188613 Mus muscu
28	621	31.8	1864	10	AF202076 Rattus no
29	621	31.8	1892	10	BC061800 Rattus no
30	620.5	31.8	185788	10	AC124461 Mus muscu
31	620.5	31.8	219763	10	AC093175 Mus muscu
32	620	31.7	2208	10	AB017638 Rattus no
33	613	31.4	1445	10	AY266139 Mus muscu
34	612	31.3	1161	6	AX675581 Sequence
35	611	31.3	244161	2	AC117170 Rattus no
36	609	31.2	1360	10	AY262280 Mus muscu
37	608	31.1	1488	10	BC024903 Mus muscu
38	603.5	30.9	1606	6	AX676264 Sequence
39	603.5	30.9	1613	6	AR256990 Sequence
40	603.5	30.9	1613	9	AF536382 Homo sapi
41	603.5	30.9	1718	9	BC062334 Homo sapi
42	603.5	30.9	1755	9	BC036846 Homo sapi
43	597	30.6	944	6	AX375744 Sequence
44	591	30.3	1020	6	AX823252 Sequence
45	590.5	30.2	1458	5	BC056000 Xenopus 1

ALIGNMENTS

RESULT 1

US-10-037-417-46 (1-357) x AX512289 (1-1102)

Qy	1	MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer	20
Db	19	ATGGGCCCAAGAGGGGGTCTCTGGGGGCTGGGACACTGGGGGCTGTGGCCAAATTCGACTCA	78

```
QY 21 TysSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProTyrCysGlyArg 40
Db 79 TACTCACTTACGGGTGGTGGCGTCCGACCCGCTAGGAGCCCCCGCTACCTGCGGGCGC 138
QY 41 ProGluProSerAlaArgLeuValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
Db 139 CTTGAGCCCTCGGCCGCTCGTGGGGGCTCAACCGCGCAGCGCGGACCTGGCCTTGG 198
QY 61 GlnValSerLeuHisHisGlyGlyHisHisCysGlyGlySerLeuLeuAlaProSer 80
Db 199 CRAAGTGAAGCTGCACCATGAGGTGGCCACATCTGCGGGGCTCCCTCATCGCCCCCTCC 258
QY 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
Db 259 TGGGTCTCTCCCGCTGCTCTGTTTCATGACGAAATGGAGCTGGAGCCCGCGCGAG 318
QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
Db 319 TGGTCTGCTACTGCTGGCGGTGCACTCCACGAGCGGCCCTGGACGGCGGCACACCCGC 378
QY 121 AlaValAlaAlaLeuValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
Db 379 GCAGTGGCGGCCCATCGTGGTGGCGGCCCACTACAGCCCAAGTGGAGCTGGCGCGGACCTG 438
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160
Db 439 GCCTCTGCTGGCGCTGGCTCACCAGCAGCGCTGGGCCCGCCGCTGGCTCTGCTGCTG 498
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
Db 499 CCCCAGCCCTCACACCGCTTGGTGGCACGCGACCCGCTGCTGGGCCACCGCGCTGGGAGAC 558
QY 181 ValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnValGluLeuArgLeu 200
Db 559 GTCCAGAGGAGCAGATCCTCTGCTCTCCCTGGGTGCTACAGGAAGTGGAGCTAAGGCTG 618
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
Db 619 CTGGCGAGGAGCCACTGCTCAATGTCTCTACAGCAGCGCCCGTCCCTCAACCTCACTCTC 678
QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240
Db 679 CAGATATTGCCAGGATGCTGTGTGTGGCTTACCAGAGGGCGCGCAGGACACCTGCCAG 738
QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIle 260
Db 739 GGTGACTCTGGGGGGCCCCCTGGTCTGTGAGAGAGCGCGCGCTGGTTTCCAGGAGGAATC 798
QY 261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
Db 799 ACCAGCTTTGGGTTTGGCTGTGGACGAGAGAAACCGCCCTGGAGTTTCACTGCTGGCT 858
QY 281 ThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyProAlaPhePro 300
Db 859 ACCTATGAGGCAATGGATACGGGAGCAGGTGATGGGTTTACAGACCTGGGCGCTTTCCC 918
QY 301 ThrGlnProGlnThrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320
Db 919 ACCAGCCCCAGAGAGCCAGTCAAGTTGTTTATCAAAAGCGCATTCCTGGATTCGCTGCC 978
QY 321 ArgIleLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrLysSer 340
Db 979 AGAATCTTTTGGGCCCTTGTCCCATATATCAGTAGGAGTCTCACTGGGACCAAAAGC 1038
QY 341 LeuValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db 1039 CTTGTCTCTCCCTGGCTCTCTCCACACTCTCTCTGGGCGCTCTGGGGGTTTC 1089
```

RESULT 3

AX360096

LOCUS

DEFINITION

Sequence 52 from Patent W00200860.

2457 bp

DNA

linear

PAT 13-FEB-2002

```
ACCESSION AX360096
VERSION AX360096.1 GI:18675722
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Plozman,G., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and Charydzak,G.
TITLE Novel proteases
JOURNAL Patent: WO 0200860-A 52 03-JAN-2002;
Sugen, Inc. (US)
FEATURES
    Location/Qualifiers
        1..2457
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 5,28e-73 Length: 2457
Score: 1505.00 Matches: 272
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 77.06% Indels: 0
DB: 6 Gaps: 0
US-10-037-417-46 (1-357) x AX360096 (1-2457)
QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
Db 112 TGCGGCGCGCCCTGAGCCCTCGGCCGCTCGTGGGGGGCTCAAAACGCGCAGCGCGGCACC 171
QY 58 TrpProTrpGlnValSerLeuHisHisGlyGlyHisIleCysGlySerLeuIle 77
Db 172 TGGCCCTTGGCAAGTGAAGCTGACCATGGAGGTGGCCACATCTCGCGGGGCTCCCTCATC 231
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
Db 232 GCCCCCTCTGGGTCTCTCCGCTGCTCACTGTTTCAATGACGAATGGAGCTGGAGCCC 291
QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
Db 292 GCGGCGCAGTGGTGGTACTGCTGGCGCTGCTCCAGGACGGCGCCCTGGAGCGCGCG 351
QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGly 137
Db 352 CACACCGCGCAGTGGCGCCATCGTGGTGGCGGCCCACTACAGCCAAAGTGGAGCTGGGC 411
QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
Db 412 GCCGACTGGCCCTGCTGGCGCTGGCCCTCACCGCCAGCGCTGGCGCCCGCTGGCCCT 471
QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
Db 472 GTCTGCTGTCGGCGCGCTCACACCGCTTGTGTCAGCGCACCGCCCTGCTGGCGCACCGGC 531
QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
Db 532 TGGGGAGACGCTCAGAGGAGCAGATCCTCTCCCTCTCCCTGGGTGCTTACAGGAAGTGGAG 591
QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
Db 592 CTAAAGCTGTGGCGGAGGCCACCTGTCTATGTCTCTACAGCCAGCCCGCTCCCTCAAC 651
QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
Db 652 CTCACTCTCCAGATATTGCCAGGAGATGCTGTGTGCTGGCTACCCAGAGGCGCGCAGGAC 711
QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGlyArgTrpPheGln 257
Db 712 ACCTGCCAGGGTGACTCTTGGGGGGCCCCCTGTGTGTGAGGAAGCGCGGCCCTGTGTTCCAG 771
```


ORIGIN	AK075142	GI:22761040	oigo capping; fis (full insert sequence).	AK075142	GI:22761040	oigo capping; fis (full insert sequence).
Alignment Scores:	3.02e-72	Length: 2810		Alignment Scores:	3.02e-72	Length: 2810
Pred. No.:	1492.00	Matches: 272		Pred. No.:	1492.00	Matches: 272
Score:	99.63%	Conservative: 0		Score:	99.63%	Conservative: 0
Best Local Similarity:	99.63%	Mismatches: 1		Best Local Similarity:	99.63%	Mismatches: 1
Query Match:	76.40%	Indels: 0		Query Match:	76.40%	Indels: 1
DB:	6	Gaps: 0		DB:	9	Gaps: 0
US-10-037-417-46 (1-357) x BD127529 (1-2810)				US-10-037-417-46 (1-357) x AK075142 (1-2810)		
QY	38	CysGlyArgProGluProSerAlaArgIleVal-GlyGlySerAsnAlaGlnProGlyTh	57	QY	38	CysGlyArgProGluProSerAlaArgIleVal-GlyGlySerAsnAlaGlnProGlyTh
DB	143	TGCGGGCGCCCTGAGCCCTTCGGCCCGCATCTGCGGGGGGCTCAACGCGCAGCCGCGGCAC	202	DB	143	TGCGGGCGCCCTGAGCCCTTCGGCCCGCATCTGCGGGGGGCTCAACGCGCAGCCGCGGCAC
QY	57	rTTPProTrpGlnValSerLeuHisHisGlyGlyHisIleCysGlyGlySerIleuI	77	QY	57	rTTPProTrpGlnValSerLeuHisHisGlyGlyHisIleCysGlyGlySerIleuI
DB	203	CTGGCCTTGCAAGTGAGCTGCGCATGAGCGAGTGGCCACATCTGCGGGGGCTCCCTCAT	262	DB	203	CTGGCCTTGCAAGTGAGCTGCGCATGAGCGAGTGGCCACATCTGCGGGGGCTCCCTCAT
QY	77	eAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPr	97	QY	77	eAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPr
DB	263	CGCCCCCTCTGGGTCTCTCCGCTCTCACGTTCATGACGAATGGGACCGCTGGAGCC	322	DB	263	CGCCCCCTCTGGGTCTCTCCGCTCTCACGTTCATGACGAATGGGACCGCTGGAGCC
QY	97	oAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAl	117	QY	97	oAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAl
DB	323	CGCGCGGAGTGTCTGCTGCTGGGGTGTCTCTCCAGGAGCGGCCCTCGAGCGGCGC	382	DB	323	CGCGCGGAGTGTCTGCTGCTGGGGTGTCTCTCCAGGAGCGGCCCTCGAGCGGCGC
QY	117	aHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuG1	137	QY	117	aHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuG1
DB	383	GCACACCCGCGCAGTGGCCCATCTGCTGCTGGCGCCCACTACAGCCAACTGGAGCTGGG	442	DB	383	GCACACCCGCGCAGTGGCCCATCTGCTGCTGGCGCCCACTACAGCCAACTGGAGCTGGG
QY	137	yAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPr	157	QY	137	yAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPr
DB	443	CGCGGACCTGGCCCTGCTGGCGCTGGCCCTCACCGCGCAGCTGGGGCCCCCGCTGGCC	502	DB	443	CGCGGACCTGGCCCTGCTGGCGCTGGCCCTCACCGCGCAGCTGGGGCCCCCGCTGGCC
QY	157	oValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrG1	177	QY	157	oValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrG1
DB	503	TGCTTGCTGCTGCGCCGCGCTCTACACCGCTTCTGTCAGCGGCACCGCTGCTGGCCACCGG	562	DB	503	TGCTTGCTGCTGCGCCGCGCTCTACACCGCTTCTGTCAGCGGCACCGCTGCTGGCCACCGG
QY	177	yTTPGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValG1	197	QY	177	yTTPGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValG1
DB	563	CTGGGGAGAGCTCCAGGAGGAGATCTCTGCTCTCTCCCTCTGCTGCTGCTGCTGCTGCTG	622	DB	563	CTGGGGAGAGCTCCAGGAGGAGATCTCTGCTCTCTCCCTCTGCTGCTGCTGCTGCTGCTG
QY	197	uLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAs	217	QY	197	uLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAs
DB	623	GCTAAGGCTGCTGGCGGAGGCGCACCTGCTCAATGCTCTACAGCCGCGCGCTCCCTCAA	682	DB	623	GCTAAGGCTGCTGGCGGAGGCGCACCTGCTCAATGCTCTACAGCCGCGCGCTCCCTCAA
QY	217	nLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAs	237	QY	217	nLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAs
DB	683	CCTCACTCTCCAGATATTCCAGGATGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	742	DB	683	CCTCACTCTCCAGATATTCCAGGATGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
QY	237	pThrCysGlnGlyAspSerGlyProLeuValCysGlnGluGlyArgTrpPheG1	257	QY	237	pThrCysGlnGlyAspSerGlyProLeuValCysGlnGluGlyArgTrpPheG1
DB	743	CACCTGCCAGGCGCCTCTGGGGGGCCCTGCTGTGAGAGGCGCGCTGCTGCTGCTGCTGCTG	802	DB	743	CACCTGCCAGGCGCCTCTGGGGGGCCCTGCTGTGAGAGGCGCGCTGCTGCTGCTGCTGCTG
QY	257	nAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheTh	277	QY	257	nAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheTh
DB	803	GGCAGGATATCCAGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	862	DB	803	GGCAGGATATCCAGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
QY	277	rAlaValAlaThrTyrGluAlaTrpIleArgGlnGlnValMetGlySerGluProGlyPr	297	QY	277	rAlaValAlaThrTyrGluAlaTrpIleArgGlnGlnValMetGlySerGluProGlyPr
DB	863	TGCTGTGCTACCTATGAGGATGATACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	922	DB	863	TGCTGTGCTACCTATGAGGATGATACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY	297	oAlaPheProThrGlnProGlnLysThrGlnSerAsp	309	QY	297	oAlaPheProThrGlnProGlnLysThrGlnSerAsp
DB	923	TGCTTTCCTCCACCCAGCCCGCCAGAGACCCAGTCAGAT	959	DB	923	TGCTTTCCTCCACCCAGCCCGCCAGAGACCCAGTCAGAT
RESULT 6				RESULT 6		
AK075142				AK075142		
LOCUS				LOCUS		
DEFINITION				DEFINITION		
Homo sapiens cDNA FLJ290661 fis, clone PLACE1005003, weakly similar				Homo sapiens cDNA FLJ290661 fis, clone PLACE1005003, weakly similar		
to PROSTASIN PRECURSOR (EC 3.4.21.-).				to PROSTASIN PRECURSOR (EC 3.4.21.-).		

COMMENT	OS	Homo sapiens (human)
PN	JP	2002017375-A/650
PD	22-JAN-2002	
PF	07-JUL-2000	JP 20000253172
PI	TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO	
PI	I-SHII,	
PI	YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI	

RESULT 9	BD126485	BD126485	670 bp	DNA	linear	PAT 18-SEP-2002
LOCUS	BD126485	BD126485				
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.					
ACCESSION	BD126485					
VERSION	BD126485.1	GI:23221430				
KEYWORDS	UP 2002017375-A/1916.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					

DEFINITION Homo sapiens chromosome 16 clone RP11-388M20, complete sequence.
AC009088
AC009088.9 GI:29366934
VERSION
HTG.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 127769)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 127769)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (03-APR-1999) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 127769)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-APR-2002) Production Sequencing Facility, DOE Joint
REFERENCE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS 4 (bases 1 to 127769)
REFERENCE DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
REFERENCE Drive, Walnut Creek, CA 94598, USA
AUTHORS 5 (bases 1 to 127769)
REFERENCE DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
AUTHORS Drive, Walnut Creek, CA 94598, USA
TITLE On Mar 29, 2003 this sequence version replaced gi:29029216.
JOURNAL Draft Sequence Produced by DOE Joint Genome Institute
COMMENT www.jgi.doe.gov
Location/Qualifiers
1. 127769
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-388M20"

FEATURES
source
1. 127769
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-388M20"

ORIGIN
Alignment Scores: 1.04e-28 Length: 127769
Pred. No.: 716.50 Matches: 155
Score: 50.81% Conservative: 1
Percent Similarity: 50.49% Mismatches: 5
Best Local Similarity: 36.69% Indels: 146
Query Match: 4
DB: Gaps: 4

US-10-037-417-46 (1-357) x AC009088 (1-127769)
QY 38 CysGlyArgProGluProSerAlaArgIleValGlyCysSerAsnAlaGlnProGlyThr 57
Db 32148 TGGGGGGCGCCCTGAGCCCTCGCCCGCATCTGGGGGGCTCAAACGCGACGGGGCACC 32089
QY 58 TrpProTrpGlnValSerLeuHisGlyGlyHisIleCysGlyGlySerLeuIle 77
Db 32088 TGGCCTTGGCAAGTGAGCCTGCACATGGAGGTGGCCATCTCGGGGGGCTCCCTCATC 32029

RESULT 10
LOCUS AX342936/c
DEFINITION Sequence 3 from Patent WO0198467.
AC009088
AC009088.1 GI:18152214
VERSION
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Xiao, Y. and Morozov, V.
TITLE Regulation of human prostasin-like serine protease
JOURNAL Patent: WO 0198467-A 3 27-DEC-2001;
Bayer Aktiengesellschaft (DE)
FEATURES
source
1. 456
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores: 2.04e-35 Length: 456
Pred. No.: 799.00 Matches: 147
Score: 97.37% Conservative: 1
Percent Similarity: 96.71% Mismatches: 4
Best Local Similarity: 40.91% Indels: 0
Query Match: 6
DB: Gaps: 0

US-10-037-417-46 (1-357) x AX342936 (1-456)
QY 97 ProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGly 116
Db 456 CCGCGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 397
QY 117 AlaHisThrArgAlaValAlaAlaIleValProAlaAsnTrpSerGlnValGluLeu 136
Db 396 GCGACACCCGCGAGTGGCGGCACTGCTGGCGGCACTACAGCCAGTGGAGCTG 337
QY 137 GlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrp 156
Db 336 GCGCGCGAGTGGCGCTGGTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGG 277
QY 157 ProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThr 176
Db 276 CCTGTCTGCTGGCGCGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGG 217
QY 177 GlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluVal 196
Db 216 GCCTGGGGAGAGCTGCCAGGAGAGATCTCTGCTCTCTCCCTGGGTGCTACAGGAAGT 157
QY 197 GluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTrpSerGlnProGlyProPhe 216
Db 156 GAGCTAAGCTCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 97
QY 217 AsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArg 236
Db 96 AACCTCACTCTCCAGATATGGCAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 37
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeuVal 248
Db 36 GACACTGCGCAGGGTGACTCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1

RESULT 11
LOCUS AC009088/c
LOCUS AC009088 127769 bp DNA linear PRI 29-MAR-2003

QY	78	AlaProSerTrpValIleuSerAlaAlaHisCysPheMetThr-----	91
DB	32028	GCCTCCCTGGGTCTCTCCGTGCTCACTGTTTCATGACGTGAGTATTCGCCCTGTGTC	31969
QY	91	-----	91
DB	31968	CTGGCCCTCTCCGCCAAAGAGCTGGCCCCCGCCCGCCAGTTGGTTCCTCTCTCCGCC	31909
QY	91	-----	91
DB	31908	CAC TTCGTCTCTCAAGGATAATTGGAGCCAAAGCTGTCCCAAGCGAGGTGCTATTGCA	31849
QY	91	-----	91
DB	31848	GGCATCATGTTCCGAGTGGGGCAAGGCTGCCGACTGGTCCACCACTCCACTCAATAT	31789
QY	91	-----	91
DB	31788	CCGGTCTGAATCCCATCTTTTCCAATGACCCCCCAACCAAGCTCCAAATCGGATTTC	31729
QY	91	-----	91
DB	31728	TCTTCCCTTTGAGAGTATCAGCACCAAATCTGGATCGATAGCACCTGCCCTCTGGC	31669
QY	91	-----	91
DB	31668	TCTGAGCCCTGCCCCCAACGAGGCTGTGCCCTCCATAGGCTCAACTAACCTCCTCTAC	31609
QY	92	-----AsnGlyThrIleuGluProAlaAlaGluTrpSerValIleuLe	105
DB	31608	TGCCCCGGTTCGAGCGCAGAAATGGGACGCTGGAGCCCGGCCGAGTGGTTCGCTACTGCT	31549
QY	105	uGlyValHisSerGlnAspGlyProIleuAspGlyAlaHisThrArgAlaValAlaAlaI	125
DB	31548	GGCGTGCATCTCCAGSACGGGCCCTCTGGACGGCGGCACACCCGCGCAGTGGCGCCAT	31489
QY	125	eValValProAlaAsnTyrSerGlnValGluIleuGlyValaAspLeuAlaIleuIleuArgLe	145
DB	31488	CGTGGTTCGGCCCAACTACAGCCAAAGTGGAGCTGGGCGCCGACCTTGGCCCTGTGCGCT	31429
QY	145	uAlaSerProAlaSerIleuGlyProAlaValTrpProValCysIleuProArgAlaSerHi	165
DB	31428	GGCTTCACCGCCGACCTGGGCCCGCCGCTGTGGCTGTCTGCTGCCCGCCGCTCACA	31369
QY	165	sArgPheValHisGlyThrAlaCysTrpAlaThrGlyTroGlyAspValGlnGluAlaAs	185
DB	31368	CCGCTTCGTGCACGGCACCGCTCTGGCCACCGGCTGGGGAGACGTCACAGGAGGCAGG	31309
QY	185	pProIleuProIleuProTrp-----ValIleuGlnGluVal	196
DB	31308	T-----GAGTGGGACGGACAGAGGGAGGCGCTCCGGGAGGTGGGCTTCG	31261
QY	197	GluLeuArgLeuLeuGly	202
DB	31260	AGAGGTGAGTTCTTAGGG	31242

RESULT	12
AX098193	
LOCUS	AX098193
DEFINITION	Sequence 105 from Patent WO0118542.
ACCESSION	AX098193
VERSION	AX098193.1 GI:13515276
KEYWORDS	.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 Lee,J., Thompson,P. and Lillie,J.
AUTHORS	Identification, assessment, prevention, and therapy of ovarian
TITLE	cancer
JOURNAL	Patent: WO 0118542-A 105 15-MAR-2001; Millennium Predictive Medicine, Inc. (US)

FEATURES
source

ORIGIN

Alignment Scores:
Pred. No.: 2,4e-29 Length: 1796
Score: 697.00 Matches: 160
Percent Similarity: 55.29% Conservative: 49
Best Local Similarity: 42.33% Mismatches: 133
Query Match: 35.69% Indels: 36
DB: 6 Gaps: 11

US-10-037-417-46 (1-357) x AX098193 (1-1796)

QY 1 MetAlaGlnIysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
DB 152 ATGGCCACGAAGGGGGTCTCTGGGCTTGGGAGCTGGGGGCTGTGGCC----- 199
QY 21 TyrSerLeuTyr-----GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
DB 200 ATTCTGCTCTATTCTTGGATTACTTCGGTCTGGGGACAGAGCGGAAGGGGCAGAGCTCCC 259
QY 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
DB 260 -----TGGCGT---GTGGCCCCCAGCACGCATCACAGGTGGCAGCAGTGCAGTCGCC 310
QY 56 GlyThrTrpProTglnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
DB 311 GGTCAGTGGCCCTGGCAGGTTCAGCATCCTATGAAGGCGTCATGTGTGTGGTGGCTCT 370
QY 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
DB 371 CTCGTGTCTGAGCAGTGGGTGTGTCTGAGTGTCTACTGTCTTCCCAGCAGCACCAAG 430
QY 96 GlnProAlaAlaGlnTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
DB 431 GAA-----GCCTATGAGGTCAAGCTGGGGGCCACCAGCTAGACTCCTACTCCGAG 481
QY 116 GlyAlaHisThrArgAlaValAlaIleValProAlaAsnTyrSerGlnValGlu 135
DB 482 GACGCCAAGGTTCAGCACCCCTGAGGACATCATCCCCACCAGCTACCTCTCCAGAGGGC 541
QY 136 LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
DB 542 TCCAGGGCGACATGTGACATCTCCAACTCAGCAGACCCATCCTTCTCCCGTACATC 603
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
DB 602 CGGCCCATCTGCCTCCCTGCGAGCCACAGCCCTCTTCCCCAACGGCCCTCCACTGCATGTC 663
QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGlu 191
DB 662 ACTGGTGGGGTCAATGTGGCCCTCTCAGTAGCCCTCTCAGCCCAAGCCACTGCAGCAA 721
QY 196 ValGlnLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 211
DB 722 CTCGAGTGGCCTTGATCTGCTGAGACGCTGTAACCTGCTGTACACATCAGCCCAAG 781
QY 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 231
DB 782 CTGAGGAGCCGCACTTGTCCAAAGAGACATGGTGTGTGTGCTATGTGAGGGGGGCG 841
QY 236 ArgAspThrCysGlnGlyAspSerGlyClyProLeuValCysGluGluGlyClyArgTrp 251
DB 842 AAGGACGCGCTGCCAGGGTGAATCTCTGGGGGGCCACTCTCTCGCCCTGTGAGGGTCTCTGG 901
QY 256 PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 271
DB 902 TACCTGACGGGCATTTGTGAGCTTGGGGAGATGCTCTGTGGGGCCGCAACAGGGCTGTGTG 961


```

Db      919  AAGAGCGCCTGCCAGGCTGACTCTGGGGGCCCACTCTCCCTGCGCTGTGGAGGCTCTCTGG 978
      256  PheGlnAlaGlyLeuThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
      979  TACCTGACGGGCATTTGACCTGGGGAGATGCCCTGTGGGGCCCGCAACAGGCTGTGTGTG 1038
      276  PheThrAlaValAlaThrTyrGluAlaThrPheGluGluGlnValMetGlySerGluPro 295
      1039  TACACTCTGGCTCCAGCTATGCTCTCTGGATCCAAAGCAAGGTG-----ACAGAACTC 1092
      296  GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
      1093  CAGCCTCGTGTGGTGGCCCAACACAGGAGTCCAGGCCCGACAGCAACCTCTGTGGCAGC 1152
      313  GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
      1153  CACCTGGCCTTCAGCTCTGCCCGCAGCCAGGGCTGTCTGAGGGCCCACTCTTTCTCTGCT 1212
      332  ValGlyValSerThrGlyThrLysSerLeuValLeuProThrLeuSer----- 347
      1213  CTGGGCTGTGGTCTGGGC-----CTGCTCTCCCATGCTCAGCGAGCACTGAGCT 1263
      348  -----ProHisSerLeuLeuGlyLeuThrGlyPhe 357
      1264  GGCCCTACTTCCAGGATGATGATCATCACTCAAGGACAGGAGCTGTGCTCTTC 1317

RESULT 15
AX336076 1834 bp DNA linear PAT 09-JAN-2002
LOCUS AX336076 Sequence 6585 from Patent WO0194629.
DEFINITION AX336076
ACCESSION AX336076
VERSION AX336076.1 GI:18126795
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 6585 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..1834
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 2,45e-29 Length: 1834
Score: 697.00 Matches: 160
Percent Similarity: 55.29% Conservative: 49
Best local Similarity: 42.33% Mismatches: 133
Query Match: 35.69% Indels: 36
Gaps: 11

US-10-037-417-46 (1-357) x AX336076 (1-1834)
      1  MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
      229  ATGGCCCAAGAGGGGTCTCTGGGGCTGGGAGCTGGGGGCTGTGGCC----- 276
      21  TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
      277  ATTCTGCTCTAATCTTGATTAATCTCCGGTCTGGGACAGGAGCGGAGGGGCGAGAGCTCCC 336
      36  ProTyrCysGlyArgProGluProSerAlaArgIleValGlySerAsnAlaGlnPro 55
      337  -----TGCGGT---GTGGGCCCCCAAGCAGCATCACAGGTGGCAGCTGCGTCGCC 387

```

Search completed: February '25, 2004, 17:20:01
Job time : 3826 secs

```

QY      56  GlyThrTrpProThrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
      388  GGTCAAGTGGCCCTGGCAGGTCACTTATGAAGGCGTCACTGTGTGTGTGTGTGTGTGTGT 447
      76  LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
      448  CTGCTGTCTGAGCAGTGGGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 507
      96  GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
      508  GAA-----GCCTATGAGTCAAGCTGGGGGCCCAACAGCTGAGCTCTCTCTCTCTCTCT 558
      116  GlyAlaHisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGlu 135
      559  GAGGCCAAGCTCAGCAACCTTGAAGGACATATCCCCACCCAGCTACCTCCAGGAGGGC 618
      136  LeuGlyAlaAspLeuAlaLeuLeuArgLeuLeuAspProAlaSerLeuGlyProAlaVal 155
      619  TCCAGGGCGCAGATGACCTCTCTCAACTCAGCAGACCCATCACCTTCTCCCGCTACATC 678
      156  TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
      679  CGGCCCATCTGCTCTCTCTCAGCCCAACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 738
      176  ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
      739  ACTGGCTGGGTCTGTGGCCCTCTAGTGAGCTCTCTGAGCCCTCTGAGCCCAAGCCACTGC 798
      196  ValGlnLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
      799  CTCGAGGTGCTCTGTATCAGTCAGTCGTGAGACGTGTAACTGCTGTATCAACATCGAC 858
      216  PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
      859  CCTGAGGAGCGCACTTTGTCCAGAGGACATGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 918
      236  ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGlyArgTrp 255
      919  AAGGAGCGCTGCCAGGTGACCTCTGGGGGCCCACTCTCTCTCTCTCTCTCTCTCTCTCT 978
      256  PheGlnAlaGlyIleThrSerPheGlyCysGlyArgArgAsnArgProGlyVal 275
      979  TACCTGACGGGCATTTGAGCTGGGAGATGCTGTGTGGGGCCCGCAACAGGCTGTGTGTG 1038
      276  PheThrAlaValAlaThrTyrGluAlaThrPheGluGlnValMetGlySerGluPro 295
      1039  TACACTCTGGCTCCAGCTATGCTCTCTGGATCCAAAGCAAGGTG-----ACAGAACTC 1092
      296  GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
      1093  CAGCCTCGTGTGGTGGCCCAACACAGGAGTCCAGGCCCGACAGCAACCTCTGTGGCAGC 1152
      313  GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
      1153  CACCTGGCCTTCAGCTCTGCCCGCAGCCAGGGCTGTCTGAGGGCCCACTCTTTCTCTGCT 1212
      332  ValGlyValSerThrGlyThrLysSerLeuValLeuProThrLeuSer----- 347
      1213  CTGGGCTGTGGTCTGGGC-----CTGCTCTCCCATGCTCAGCGAGCACTGAGCT 1263
      348  -----ProHisSerLeuLeuGlyLeuThrGlyPhe 357
      1264  GGCCCTACTTCCAGGATGATGATCATCACTCAAGGACAGGAGCTGTGCTCTTC 1317

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 15:50:34 ; Search time 417 Seconds
(without alignments)
3636.947 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953
Sequence: 1 MAQKGVLPGLGAVANSDS.....TKSLVLPWLSPHSLGLWGF 357

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgm2_1/USPTO.spool_p/US10037417/runat_24022004_101037_17437/app_query.fasta_1.519
-DB=N_Geneseq_29Jan04 -OPMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417_@CGN_1_1_885_@runat_24022004_101037_17437 -NCFU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04: *
1: geneseq1980s: *
2: geneseq1990s: *
3: geneseq2000s: *
4: geneseq2001as: *
5: geneseq2001bs: *
6: geneseq2002as: *
7: geneseq2003as: *
8: geneseq2003bs: *
9: geneseq2003cs: *
10: geneseq2004as: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1953	100.0	1102	6 ABQ93901	Abq93901 Human pro
2	1953	100.0	1102	6 ABQ93902	Abq93902 Human pro
3	1505	77.1	2457	6 ABK31795	Abk31795 DNA encod
4	1505	77.1	2681	6 ABQ75956	Abq75956 Human PMM
5	1492	76.4	2810	4 AAK94500	Aak94500 Human ful
6	970	49.7	537	6 ABK13565	Abk13565 Human pro
7	861	44.1	670	4 AAK92190	Aak92190 Human CDN
8	861	44.1	670	4 AAK93456	Aak93456 Human CDN

c	9	799	40.9	456	6	ABK13566	Human pro
	10	787.5	40.3	597	5	AAS69040	DNA encod
	11	697	35.7	1733	7	ABT31936	Human bre
	12	697	35.7	1796	5	Aaf98698	Human ova
	13	697	35.7	1834	6	ABF76501	cDNA enco
	14	697	35.7	1834	6	ABL67949	Ovary can
	15	697	35.7	1834	6	ABL68512	Kidney ca
	16	697	35.7	1834	6	ABL68248	Kidney ca
	17	697	35.7	1834	6	ABK12241	cDNA enco
	18	697	35.7	1834	6	ABN95716	Gene #221
	19	697	35.7	1835	5	AAF98720	Human lat
	20	697	35.7	3382	6	ABZ35336	Human gen
	21	665	34.1	1668	3	AACT7814	Human can
	22	652	33.4	1726	6	ABN85392	Human NOV
	23	627.5	32.1	1797	7	ABN85393	Mouse NOV
	24	612	31.3	1161	6	ABN85393	Human NOV
	25	603.5	30.9	1606	6	ABN85392	DNA encod
	26	603.5	30.9	1613	4	AAO1360	Human ser
	27	597	30.6	944	6	ABA94396	Human pro
	28	596.5	30.5	849	6	ABQ86176	Novel hum
	29	590	30.2	843	6	ABQ86175	Novel hum
	30	589	30.2	2662	6	AAD23854	Human pro
	31	585.5	30.0	2122	5	AAS74051	DNA encod
	32	582.5	29.8	1958	6	ABK12891	Human pro
	33	576	29.5	786	6	ABA94394	Human pro
	34	575.5	29.5	768	6	ABN85394	Nucleotid
	35	575.5	29.5	882	6	ABN85395	Partial H
	36	571.5	29.3	882	6	ABN85395	Partial H
	37	567.5	29.1	1703	3	AAZ52473	HTRM clon
	38	565	28.9	1887	6	ABK11770	DNA encod
	39	565	28.9	1973	7	ABZ58499	Transmemb
	40	564.5	28.9	1142	3	AAC87796	Activatio
	41	564.5	28.9	1142	4	AAF5268	Nucleotid
	42	563.5	28.9	1110	4	AAF76994	Human pro
	43	563.5	28.9	1129	4	AAS21354	Human CDN
	44	563.5	28.9	1129	7	ACD23963	Novel hum
	45	563.5	28.9	1129	7	ACA67104	cDNA enco

RESULT 1

ABQ93901

ID

ABQ93901 standard; DNA; 1102 BP.

XX

XX

XX

AC

ABQ93901;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

ALIGNMENTS

RESULT 1
ABQ93901
ID ABQ93901 standard; DNA; 1102 BP.

XX AC ABQ93901;
XX DT 01-NOV-2002 (first entry)
XX DE Human prostatin precursor-like NOV14a DNA, SEQ ID NO:43.
XX KW Human; NOX; neurological disorder; Alzheimer's disease;
XX KW Huntington's disease; Parkinson's disease; pain; behavioural disorder;
XX KW addition; tuberculous sclerosis; cancer; immune disorder; allergy;
XX KW autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;
XX KW thyroiditis; cardiovascular disease; hypertension; reproductive disorder;
XX KW endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;
XX KW pancreatitis; cirrhosis; glomerular endoheiosis; bacterial infection;
XX KW polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
XX KW atherosclerosis; cell signal processing-related disorder;
XX KW metabolic pathway regulation disorder; cytostatic; neuroprotective;
XX KW antiinflammatory; immunosuppressive; analgesic; antithrombotic;
XX KW dermatological; antibacterial; antiarthritic; hepatotrophic; neurogenesis;
XX KW differentiation; proliferation; motility; haematopoiesis; wound healing;
XX KW angiogenesis; forensic biology; transgenic animal; drug screening;
XX KW gene therapy; NOV14a; prostatic precursor-like; chromosome 16; gene; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 1..18
XX FT /*tag= a
XX FT 19..1050
XX FT CDS

PT FT /*tag= b
PT FT /partial
PT FT /product= "NOV14a"
XX XX /note= "No stop codon given in the specification"

WO200253742-A2.

11-JUL-2002.

07-JAN-2002; 2002WO-US000375.

05-JAN-2001; 2001US-0260018P.

08-JAN-2001; 2001US-0260360P.

28-FEB-2001; 2001US-0272411P.

02-MAR-2001; 2001US-0272817P.

05-JUL-2001; 2001US-0303231P.

12-JUL-2001; 2001US-0305060P.

10-SEP-2001; 2001US-0318405P.

12-SEP-2001; 2001US-0318700P.

04-JAN-2002; 2002US-00037417.

(CURA-) CURAGEN CORP.

Kekuda R, Alsbrook JP, Tchernev VT, Liu X, Spytek KA;

Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CAM, Li L;

Gorman L, Edinger S, Sciore P, Ellerman K, Malyankar U;

Rothenberg M, Stone D, Boidog F, Guo X, Shenoy S, Anderson D;

Padigaru M, Taupier RJ, Miller CE, Eisen A;

WPI; 2002-583619/62.

P-PSDB; ABH09523.

Novel polypeptides and nucleic acids homologous to transmembrane receptor, thymosin, neuromodulin-like family of proteins for diagnosing, treating cancer, atherosclerosis, neurological, skin and autoimmune disorders.

Claim 9a; Page 142; 323pp; English.

The invention relates to 24 novel human proteins designated NOV1-NOV14 (AB09501-AB09524), collectively referred to as NOVX proteins, and nucleic acids encoding them (AB093879-AB093902). NOVX proteins and nucleic acids are useful in the treatment, diagnosis or prevention of NOVX-associated disorders or in the manufacture of a medicament for treating such disorders, with specific applications described for each of the 24 NOVX proteins, based on their homology to known proteins. Various disorders are associated with NOVX proteins including neurological disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases), pain, behavioural disorders, addiction, tubercular sclerosis, cancers (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma, various forms of arthritis, diabetes, thyroiditis, cardiovascular disease (e.g., hypertension), scleroderma, alopecia, ulcers, pancreatitis, cirrhosis, glomerular endoteliosis, polycystic kidney disease, endocrine disorders, obesity, bacterial infections and particularly cardiomyopathy, atherosclerosis, cell signal processing-related disorders and disorders of metabolic pathway regulation. NOVX nucleic acids and polypeptides may be used to identify cellular receptors or downstream effectors which binds to a NOVX protein, and are also useful as targets for the identification of small molecules that modulate or inhibit processes such as neurogenesis, cell differentiation, cell motility, cellular proliferation, haematopoiesis, wound healing and angiogenesis. NOVX nucleic acid sequences can be used to identify a cell or tissue type and are useful as a source of primers or probes for forensic biology and for identifying and cloning NOVX homologues in other cell types. Cells comprising NOVX nucleic acids are useful for producing non-human transgenic animals which are useful for studying the function and activity of NOVX proteins and for identifying and evaluating modulators of NOVX activity. The present sequence represents DNA encoding the prostatic precursor-like protein NOV14a. The gene encoding NOV14a is located on chromosome 16

XX

SQ Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,39e-99 Length: 1102
Score: 1953.00 Matches: 357
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-037-417-46 (1-357) x.ABQ93901 (1-1102)

QY 1 MetAlaGlnIysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 19 ATGGCCCAAGAGGGGTCCTGGGGCCCTGGGAGCTGGGGCTGTGGCCAAATTCAGACTCA 78
QY 21 TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
Db 79 TACTCACTTTACGGGTGGTGGTCCGTCAGGAGCCCGCTAGGGGGCCCCCGGTACTGTGGGGCGC 138
QY 41 ProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
Db 139 CCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAAAACGCGCAGCGGGCAGCCTGGCCTGG 198
QY 61 GlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSer 80
Db 199 CAAAGTAGCCCTGCACCATGGAGGGTGGCCACATCTGGGGGGCTCCCTCATCGCCCTCC 258
QY 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
Db 259 TGGGTCTCTCGCTGCTCACTGTGTTTCATGACCAATGGGAGCTTGGAGCCCGGCCGAG 318
QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
Db 319 TGGTCGGTACTCTGGGGCTGCACCTCCAGAGAGGGGCCCTCGAGCGGCGCGCACCCGCG 378
QY 121 AlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
Db 379 CGAGTGGCCCGCCATCGTGGTGGCGCCCACTACAGCAAGTGGAGCTGGGGCGCGACCTG 438
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160
Db 439 GCCTCTGGCTGGCTGGCTGCCTCACCGCGAGCTGGGGCCCGCGGTGGTGGCTGTGCTGCT 498
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
Db 499 CCGCGGCGCTCACACCGCTTCGTGTCAGCGGACCGCTGTGGCCACCGCTGGGGAGAC 558
QY 181 ValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeu 200
Db 559 GTCCAGAGGAGCATCTCTGCTGCTCCCTCTCCCTGGTGTACAGGAAGTGGAGCTAAGCTG 618
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
Db 619 CTGGGCGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGTCCCTTCAACCTCACTCTC 678
QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240
Db 679 CAGATATTGCCAGGGATGCTGTGCTGGCTACCCAGAGGGGCCCGCAGGACACCTGCCAG 738
QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyArgTrpPheGlnAlaGlyTle 260
Db 739 GGTGACTCTGGGGGGCCCCCTGTGTGTGAGGAAGCGCGCTGTGTTCAGGCGAAGATC 798
QY 261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
Db 799 ACCAGCTTTGGGTTTGGCTGTGGAGCGGAGAACCGCCCTGGAGTTTTCAGTGTGTGCT 858
QY 281 ThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyProAlaPhePro 300
Db 859 ACCTATGAGGCGATGATACGGGAGCGAGGTGATGGGTTCAGAGCCTGGGCTGCCCTTCCC 918
QY 301 ThrGlnProGlnLysThrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320

Db 919 ACCGAGCCCCAGAGACCCAGTTCAGATTGTTATCATCAAAAGCGGATTCCTGGATCTGCC 978
Qy 321 ArgileLeuAigProLeuSerHisHisSerValGlyValSerThrGlyThrIlySer 340
Db 979 AGAATCCCTTTGAGGCCCTTGTCCATATATACAGTAGAGTCTCAACTGGGACCAAAAGC 1038
Qy 341 LeuValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db 1039 CTTGTCCTCCCTGGCTCTCTCCACACTCTCTCTCTGGGCGCTCTGGGGGTTC 1089
RESULT 2
ID ABQ93902 standard; DNA; 1102 BP.
XX ABQ93902;
AC ABQ93902;
XX 01-NOV-2002 (first entry)
DT Human prostatic precursor-like NOV14b DNA, SEQ ID NO:45.
DE Human; NOVX; neurological disorder; Alzheimer's disease;
KW Huntington's disease; parkinson's disease; pain; behavioural disorder;
KW addition; tuberos sclerosis; cancer; immune disorder; allergy;
KW autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;
KW thyroiditis; cardiovascular disease; hypertension; reproductive disorder;
KW endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;
KW pancreatitis; cirrhosis; glomerular endotheliosis; bacterial infection;
KW polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
KW atherosclerosis; cell signal processing-related disorder;
KW metabolic pathway regulation disorder; cytostatic; neuroprotective;
KW antiinflammatory; immunosuppressive; analgesic; antiatherosclerotic;
KW dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;
KW differentiation; proliferation; motility; haematopoiesis; wound healing;
KW angiogenesis; forensic biology; transgenic animal; drug screening;
KW gene therapy; NOV14b; prostatic precursor-like; gene; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
PH 1. .18
FT 5'UTR /*tag= a
FT CDS 19. .1092
FT /*tag= b
FT 3'UTR /*product= "NOV14b"
FT 1093. .1102
FT /*tag= c
XX WO200253742-A2.
XX 11-JUL-2002.
XX 07-JAN-2002; 2002WO-US000375.
XX 05-JAN-2001; 2001US-0260018P.
XX 08-JAN-2001; 2001US-0260360P.
XX 28-FEB-2001; 2001US-0272411P.
XX 02-MAR-2001; 2001US-0272817P.
XX 05-JUL-2001; 2001US-0303231P.
XX 12-JUL-2001; 2001US-0305060P.
XX 10-SEP-2001; 2001US-0318405P.
XX 12-SEP-2001; 2001US-0318700P.
XX 04-JAN-2002; 2002US-00037417.
XX (CURA-) CURAGEN CORP.
XX Kekuda R, Alsbrook JP, Tchernev VT, Liu X, Spytek KA;
PI Patturajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CAM, Li L;
PI Gorman L, Edinger S, Sciore P, Ellerman K, Malvankar U;
PI Rothenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;
PI Padigaru M, Taupier RJ, Miller CE, Eisen A;
XX WPI; 2002-583619/62.

DR P-PSDB; ABB09524.
XX Novel polypeptides and nucleic acids homologous to transmembrane
PT receptor, thymosin, neuromodulin-like family of proteins for diagnosing,
PT treating cancer, atherosclerosis, neurological, skin and autoimmune
PT disorders.
XX Claim 9a; Page 143; 323pp; English.
XX The invention relates to 24 novel human proteins designated NOV1-NOV14
CC (ABB09501-ABB09524), collectively referred to as NOVX proteins, and
CC nucleic acids encoding them (ABQ93879-ABQ93902). NOVX proteins, and
CC nucleic acids are useful in the treatment, diagnosis or prevention of NOVX-
CC associated disorders or in the manufacture of a medicament for treating
CC such disorders, with specific applications described for each of the 24
CC NOVX proteins, based on their homology to known proteins. Various
CC disorders are associated with NOVX proteins including neurological
CC disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
CC pain, behavioural disorders, addiction, tuberos sclerosis, cancers
CC (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
CC (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,
CC various forms of arthritis, diabetes, thyroiditis, cardiovascular disease
CC (e.g., hypertension), reproductive disorders, endometriosis,
CC incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,
CC cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
CC disorders, obesity, bacterial infections and particularly cardiomyopathy,
CC atherosclerosis, cell signal processing-related disorders and disorders
CC of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
CC be used to identify cellular receptors or downstream effectors which
CC binds to a NOVX protein, and are also useful as targets for the
CC identification of small molecules that modulate or inhibit processes such
CC as neurogenesis, cell differentiation, cell motility, cellular
CC proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
CC nucleic acid sequences can be used to identify a cell or tissue type and
CC are useful as a source of primers or probes for forensic biology and for
CC identifying and cloning NOVX homologues in other cell types. Cells
CC comprising NOVX nucleic acids are useful for producing non-human
CC transgenic animals which are useful for studying the function and
CC activity of NOVX proteins and for identifying and evaluating modulators
CC of NOVX activity. The present sequence represents DNA encoding the
CC prostatic precursor-like protein NOV14b
XX Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;
SQ Alignment Scores:
Pred. No.: 1.39e-99 Length: 1102
Score: 1953.00 Matches: 357
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 6
US-10-037-417-46 (1-357) x ABQ93902 (1-1102)
Qy 1 MetAlaGlnIlyGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 19 ATGGCCCAAGAGGGGTCCTGGGGCTGGGCGAGCTGGGGGCTGTGGCCAAATCTGACTCA 78
Qy 21 TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
Db 79 TACTCATTTCAGGGTGGTGGCTCCGACCCGCTAGGGGCCCCCGTACTGCGGGGCGC 138
Qy 41 ProGluProSerAlaArgIleValGlySerAsnAlaGlnProGlyThrTyrProTyr 60
Db 139 CCGAGCCCTCGGGCCGCATCGTGGGGGGGCTCAACAGCGCGAGCCGCGACCTGGCTTGG 198
Qy 61 GlnValSerLeuHisGlyGlyHisIleCysGlyGlySerLeuIleAlaProSer 80
Db 199 CAAGTGAGCCTGCACATGGAGGTGGCCACATCTCGGGGGGCTCCCTCATCGCCCCCTCC 258
Qy 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
Db 259 TGGGTCTCTCTCGCTGCTCACTGTTTCATGACGAATGGGACGTTGGAGCCCGGCGCGAG 318

101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
119 TGGTGGTACTGCTGGGCGTGCATCCAGGACGGGCGCCCTGGACGGCGGCACACCGC 378
121 AlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
139 GCAGTGGCGCGCAFCGTGGTGGCGGCACTACAGCCAAAGTGGAGCTGGGCGCGACCTG 438
141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyrProValCysLeu 160
159 GCGCTGCTGGCGCCCTGCCTACCGCCAGCTGGCGCCCGCGCTGTGGCTGTCTGGCTG 498
161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTyrAlaThrGlyTyrGlyAsp 180
179 CCGCGCGCTCACACCGCTTGGTGCAGCGCACCGCTGCTGGGCGACCGGTGGGAGAC 558
181 ValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGluValGluLeuArgLeu 200
199 GTCCAGGAGGAGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 618
201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
219 CTGGCGGAGGACCTGCTCAATGCTCTACAGCCAGCGCGCTGCTGCTGCTGCTGCTGCTG 678
221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240
239 CAGATATTGCCAGGAGT 738
241 GlyAspSerGlyGlyProLeuValCysGluGluGlyGlyValArgTyrPheGlnAlaGlyIle 260
259 GGTGACTCTGGGCGGCGCCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 798
261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
279 ACCAGCTTGGT 858
281 ThrTyrGluAlaTyrPheArgGluGlnValMetGlySerGluProGlyProAlaPhePro 300
299 ACCTATGAGCATGATATGAGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGAT 918
301 ThrGlnProGlnTyrThrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320
319 ACCAGCGCCGAGAGACCCAGTCTGATGTTTATCATCAACAGCGCATTCCTGGATTCTGCC 978
321 ArgIleLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrIleSer 340
359 AGAATCCTTTTGGGCGCTTGTCCATATATATATATATATATATATATATATATATATATAT 1038
341 LeuValLeuProTyrPheSerProHisSerLeuLeuGlyLeuTyrPheGlyPhe 357
1039 CTTGTCT 1089

RESULT 3
ABK31795
ID ABK31795 standard; DNA; 2457 BP.
XX
AC ABK31795;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA encoding novel human protease #52.
XX Human; protease; cancer; immune-related disorder; cardiovascular disease;
KW neuronal-associated disease; metabolic disorder; inflammatory disorder;
KW nervous system disorder; sexual dysfunction; pain; mood disorder;
KW hypertension; psychotic disorder; neurological disorder; dyskinesia;
KW viral infection; human immunodeficiency virus; HIV; non-viral infection;
KW ocular disease; cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200860-A2.

03-JAN-2002.
26-JUN-2001; 2001WO-US020171.
26-JUN-2000; 2000US-0214047P.
(SUGS-) SUGEN INC.
Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
Charyczak G;
WPI; 2002-139913/19.
P-PSDB; AAU82753.
Nucleic acids encoding novel human proteases, useful for useful for
treating diseases and disorders such as cancers, immune-related diseases
and disorders, cardiovascular disease (e.g. restenosis) and inflammatory
disorders.
Claim 26; Fig 1SS-TT; 313pp; English.
The present invention relates to the isolation of novel human proteases,
and the nucleic acids encoding them. The sequences of the invention are
useful for treating diseases and disorders such as cancers (e.g. breast,
colon, lung), immune-related diseases and disorders (e.g. inflammatory
diseases and asthma), cardiovascular diseases (e.g. restenosis and
coronary thrombosis), brain or neuronal-inflammatory disorders, metabolic
disorders (e.g. diabetes, obesity), central or peripheral nervous system
diseases, rheumatoid arthritis and peoriasis), central or peripheral nervous system
diseases, migraines, pain, sexual dysfunction, mood disorders, attention
disorders, cognitive disorders, hypotension, hypertension, psychotic
disorders, neurological disorders (e.g. Alzheimer's disease, Parkinson's
disease) and dyskinesias. The nucleic acids and polypeptides are also
useful for treating viral infections caused by human immunodeficiency
virus (HIV), and non-viral infections such as ocular disease (e.g.
glaucoma) and macular degeneration. ABK31744-ABK31802 represent DNA
sequences encoding for the novel human proteases of the invention
XX Sequence 2457 BP; 377 A; 851 C; 797 G; 432 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 1-75e-74 Length: 2457
Score: 1505.00 Matches: 272
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 77.06% Indels: 0
DB: Gaps: 0
US-10-037-417-46 (1-357) x ABK31795 (1-2457)
QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
Db 112 TGGCGGCGCGCTGAGCCCTCGGCGCGCTCGTGGGCGGCTCAACCGCAGCGCGGCACC 171
QY 58 TrpProTyrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeu 77
Db 172 TGGCCCTGGCAAGTGAGCTGACCTGACCATGGAGGTGGCCACATCTCGGGGCTCCCTCATC 231
QY 78 AlaProSerTyrValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
Db 232 GCGCCCTCTCTGGTCTCTCTCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 291
QY 98 AlaAlaGluTyrSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
Db 292 GCGCGCGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 351
QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGly 137
Db 352 CACACCGCGGAGTGGCGCGCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 411
QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyrPro 157

DB 412 GCGACCTGGCCCTGCTGGCCCTGACCGCCGCGCCGCGCTGGCCCT 471
QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysThrAlaThrGly 177
DB 472 GTCGCTGCGCCGCGCTCACACCGCTGCTGTCACGCGCCGCTGCTGGGCCACCGCC 531
QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProLeuProLeuValGlu 197
DB 532 TGGGGAGAGCTCCAGAGAGCGAGATCCTCTGCTCTCCCTCGGCTGCTACAGGAGTGGAG 591
QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
DB 592 CTAAGGCTGCTGGGAGGAGCCACCTGTCAATGCTCTACAGCAGCGCGCTGCCCTCAAC 651
QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
DB 652 CTCACCTCTCCAGATATGTCAGGAGTGTGTGTGCTGCTACCCAGAGCGCCGAGGGAC 711
QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGlyGlyArgTyrPheGln 257
DB 712 ACCTGCGAGGGTGACTCTGGGGGCGCCCTGCTGTGTGAGGAGCGCGCTGTTCCAG 771
QY 258 AlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
DB 772 GCAGGATACACAGCTTTGGCTTGGCTGGAGCGGAGAACCGCTCGAGTTTCACT 831
QY 278 AlaValAlaThrTyrGluAlaThrIleArgGluGlnValMetGlySerGluProGlyPro 297
DB 832 GCTGTGGCTACCTATGAGGATGATGATGAGGAGGAGGATGATGATGATGATGATGAT 891
QY 298 AlaPheProThrGlnProGlnTyrThrGlnSerAsp 309
DB 892 GCCTTTCCACCCAGCCCGCAGAGACCCAGTCAGAT 927
RESULT 4
ABQ75956
ID ABQ75956 standard; cDNA; 2681 BP.
XX AC ABQ75956;
XX DT 17-OCT-2002 (first entry)
XX DE Human PMMM encoding sequence Incyte ID 2751509CB1.
XX KW Human; PMMM; protein modification and maintenance molecule;
XX KW anticonvulsant; neuroprotective; nootropic; cycostatic; antipsoriatic;
XX KW antiasthmatic; dermatological; antidiabetic; antiparkinsonian;
XX KW antianemic; antiinflammatory; antitumor; antianginal; cardiant;
XX KW hepatotropic; osteopathic; antiemetic; antipyretic; laxative;
XX KW antibacterial; fungicide; gastrointestinal; antidiarrheic; antiallergic;
XX KW cardiovacular; antiarteriosclerotic; hypotensive; vasotropic;
XX KW antitumor; antirheumatic; immunosuppressive; antiallergic; antihypertensive;
XX KW nephrotropic; antitumor; thymotropic; antidiarrheic; uropathic;
XX KW ophthalmological; antiparasitic; tranquilizer; vulnerary; keratolytic;
XX KW auditory; antiseborrheic; antidepressant; neuroleptic; antinfertility;
XX KW anthelmintic; protozoacide; Crohn's disease; hypertension; autoimmune;
XX KW inflammatory; anaemia; cell proliferative; developmental; epithelial;
XX KW scabies; neurological; Alzheimer's disease; reproductive;
XX KW ectopic pregnancy; gene therapy; vaccine; disorder; prostasin; gene; ss.
OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 1..2457
XX FT /*tag= a
XX FT /product= "protein modification and maintenance molecule"
XX PN WO200246383-A2.
XX PD 13-JUN-2002.
XX PF 05-DEC-2001; 2001WO-US046964.

XX 08-DEC-2000; 2000US-0254399P.
PR 21-DEC-2000; 2000US-0257803P.
PR 05-JAN-2001; 2001US-0260110P.
PR 19-JAN-2001; 2001US-0262851P.
PR 25-JAN-2001; 2001US-0264623P.
XX (INCY-) INCYTE GENOMICS INC.
PA Yue H, Azimzai Y, Kallick DA, Baughn MR, Griffin JA, Swarnakar A;
PI Lal PG, Wallia NK, Hafalia AJA, Gandhi AR, Au-Young J, Elliott VS;
PI Rankumar J, Thangavelu K, Lu Y, Warren BA, Lu DAM, Lee EA;
PI Tribouley CM, Arvizu C, Delegeane AM, Yao MG, Khan FA;
PI Sarjanwala MM;
XX WPI; 2002-519664/55.
DR P-PSDB; ABB98135.
XX New isolated Protein Modification and Maintenance polypeptides, useful
PT for diagnosis, and treatment of e.g. gastrointestinal disorders.
XX Claim 5; Page 196; 200pp; English.
XX The invention relates to an isolated Protein Modification and Maintenance
CC (PMM) polypeptide. Polypeptides of the invention may be used in the
CC diagnosis, treatment and prevention of disorders associated with
CC decreased expression or activity of PMM. These include gastrointestinal
CC disorders (e.g. Crohn's disease), cardiovascular disorders (e.g. cell
CC hypertension), autoimmune/inflammatory disorders, epithelial disorders
CC proliferative disorders, developmental disorders (e.g. Alzheimer's disease)
CC (e.g. scabies), neurological disorders (e.g. ectopic pregnancy), and in gene therapy or a
CC vaccine for such diseases. They may also be used in the assessment of the
CC effects of exogenous compound on the expression of nucleic acid and amino
CC acid sequences of protein modification and maintenance molecules. The
CC current sequence represents a human PMM encoding sequence of the
CC invention, encoding a polypeptide which has been found to have homology
CC with rat prostasin
XX SQ Sequence 2681 BP; 431 A; 919 C; 861 G; 470 T; 0 U; 0 Other;
Alignment Scores:
Pred. NO.: 1.9e-74 Length: 2681
Score: 1505.00 Matches: 272
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 77.06% Indels: 0
DB: Gaps: 6
US-10-037-417-46 (1-357) x ABQ75956 (1-2681)
QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
DB 112 TGGGGCGCCCTGAGCCCTCGCGCATCTGGGGGGCTCAACCGCGAGCGCGGCGACC 171
QY 58 TrpProTyrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIle 77
DB 172 TGGCCTTGGCAGTGGAGCTGCGCCATGAGAGTGGCCACATCTGCGGGCTCCCTCATC 231
QY 78 AlaProSerTyrValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
DB 232 GCCCCTCTCTGGGTCTCTCCGCTGCTCTCATGTCATGACGATGGGACGCTGGAGCCC 291
QY 98 AlaAlaGluTyrSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
DB 292 GCGGGCGAGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 351
QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGly 137
DB 352 CACACCGCGCAGTGGCCGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 411
QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyrPro 157

Db 863 TGCTGTGGCTACTATGAGCCATGATCGGAGCAGGATGATGGGTTCAGAGCCTGGGGCC 922

Qy 297 oAlaPheProThrGlnProGlnMysThrGlnSerAsp 309

Db 923 TGCTTTCCACCCAGCCGCCAGAGACCCAGTCAGAT 959

RESULT 6

ABK13565/c
ID ABK13565 standard; cDNA; 537 BP.

XX ABK13565;

XX 08-MAY-2002 (first entry)

XX Human prostatic-like serine protease cDNA #1.

XX Human; prostatic-like serine protease; cytosolic; antiatherosclerotic;
KW virucide; osteopathic; antiinflammatory; vasotropic; neuroprotective;
KW trypsin-like; metastasis; autoimmune lesion; atherosclerosis; gene;
KW renal failure; malignant cell; tumour angiogenesis; enzyme; inflammation;
KW chronic obstructive pulmonary disease; COPD; restenosis; atherosclerosis;
KW neurodegenerative disease; prion protein; infection; amyloid plaque;
KW Gensmann-Straussler Syndrome; viral infection; Scrapie;
KW Creutzfeldt-Jakob disease; metastatic cancer; lipid accumulation;
KW osteoporosis; Paget's disease; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
XX CDS complement(3. .536)

XX /*tag= a

XX /product= "Prostatic-like enzyme"

XX /partial

XX /note= "No start or stop codons shown"

XX W0200198467-A2.

XX 27-DEC-2001.

XX 22-JUN-2001; 2001WO-EP007117.

XX 23-JUN-2000; 2000US-0213588P.

XX 20-MAR-2001; 2001US-0276909P.

XX (FARB) BAYER AG.

XX Xiao Y, Morozov V;

XX WPI: 2002-114576/15.

XX P-PSDB; AAU75082.

XX Novel human prostatic-like serine protease polypeptide and polynucleotide
XX which can be regulated for treating metastasis of malignant cells,
XX inflammation, atherosclerosis, neurodegenerative disease and infections.

XX Claim 1; Fig 1; 111pp; English.

XX This invention comprises the cDNA and protein sequences of an isolated
XX prostatic-like serine protease and reagents and methods for regulating
XX the human prostatic-like enzyme activity. Prostatic is a trypsin-like
XX serine protease purified from human seminal fluid. An antibody specific
XX for prostatic-like serine protease is useful for immunodetection and
XX diagnosis of micro-metastases, autoimmune lesions and renal failure in
XX biopsy specimens, plasma samples and body fluids. The antibody may be
XX used to modulate enzyme activity in a disease, such as metastasis of
XX malignant cells, tumour angiogenesis, inflammation, chronic obstructive
XX pulmonary disease (COPD), atherosclerosis, neurodegenerative disease
XX (e.g. prion protein amyloid plaques of Gensmann-Straussler Syndrome,
XX Creutzfeldt-Jakob disease, Scrapie) or infection, particularly viral
XX infection. The human prostatic-like serine protease gene provides a
XX therapeutic target of decreasing human prostatic-like serine protease
XX activity, in particular for treating or preventing metastatic cancer. The
XX agonists and antagonists of the nucleotide sequence may be used to mimic,

CC augment and inhibit the enzyme activity which may be useful to treat
CC osteoporosis, Paget's disease and degradation of bone implants
CC particularly dental implants. Altered levels of human prostatic-like
CC serine protease activity inhibits both smooth muscle cell proliferation
CC and lipid accumulation and inhibit the progression of restenosis and
CC atherosclerosis. The nucleic acid sequence is also useful in diagnostic
CC assays for detecting diseases and abnormalities or susceptibility to
CC diseases related to the presence of mutations in nucleic acid sequences
CC which encode the enzyme. The present sequence represents the human
CC prostatic-like serine protease #1 nucleotide sequence of the invention
XX
XX Sequence 537 BP; 92 A; 176 C; 197 G; 72 T; 0 U; 0 Other;

Alignment Scores: Length: 537
Pred. No.: 1.48e-45 Matches: 178
Score: 970.00 Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 49.67% Gaps: 0
DB: 6

US-10-037-417-46 (1-357) x ABK13565 (1-537)

Qy 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db 536 CTATCGCCCCCTCCGGGTCTCTCGGTGCTCACTGTTTCATGACGAATGGGACGCTG 477

Qy 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 476 GAGCCCGCGCGAGTGGTGGTCTGCTGGGGCTGCACCTCCAGGACGGCGCCCTGGAC 417

Qy 116 GlyAlaHisThrArgAlaValAlaAlaValValProAlaAsnTyrSerGlnValGlu 135
Db 416 GGGCGCGCACACCGCGCAGTGGCGCCCATCTGTGTGGCCGACCTACAGCCAGTGGAG 357

Qy 136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 356 CTGGCGCGCCGACCTGGGCTGTGGGCTGGCCCTCACCGCCAGCCTGGCGCCGCGGTG 297

Qy 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 296 TGGCTCTGTCTGCTGCTGGCGCGCTCACAGCGCTTCGTGCACGGCACCGCTGTGGGCC 237

Qy 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
Db 236 ACCGGCTGGGAGACGTCAGAGGCGAGATCTCTGCTCTCCCTGGGTCTACAGGAA 177

Qy 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db 176 GTGGAGCTAAGGCTGTGGGCGAGGCCACCTGTCAATGTCTCTACAGCCAGCCGCTGCC 117

Qy 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetIleuCysAlaGlyTyrProGluGlyArg 235
Db 116 TTCAACCTCACTCCAGATATTGCCAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 57

Qy 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGly 253

Db 56 AGGGACACTGCCAGGGTGACTCTCTGGGGGGCCCTCTGGTCTGTGTGTGTGTGTGTGTGTGT 3

RESULT 7

AAK92190

ID AAK92190 standard; cDNA; 670 BP.

XX AAK92190;

XX 06-NOV-2001 (first entry)

XX Human cDNA 5'-end sequence, SEQ ID NO: 650.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

XX

PN EPI130094-A2.
XX 05-SEP-2001.
XX 07-JUL-2000; 2000EP-00114089.
XX 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI; 2001-524255/58.
XX 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX Claim 2; SEQ ID NO 650; 1380pp + Sequence Listing; English.
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is the nucleotide sequence of the 5'-end of
CC a cDNA provided in the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in CD-
CC ROM format directly from EPO
XX
SQ Sequence 670 BP; 97 A; 245 C; 205 G; 118 T; 0 U; 5 Other;
Alignment Scores:
Pred. No.: 1.94e-39 Length: 670
Score: 861.00 Matches: 167
Percent Similarity: 94.92% Conservatives: 1
Best Local Similarity: 94.35% Mismatches: 5
Query Match: 44.09% Indels: 4
DB: Gaps: 0
US-10-037-417-46 (1-357) x AAK92190 (1-670)
QY 38 CysGlyArgProGluProSerAlaArgIleVal-GlyGlySerAsnAlaGlnProGlyTh 57
DB 143 TGGGGCGCCCTGAGCCCTCGCCCGCATGCTGGGGGGGCTCAACCGCGAGCGGGCAC 202
QY 57 rTTPProTrpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeu1 77
DB 203 CTGGCCCTGGCAAGTGGAGCCCTGCACCATGAGTGGGACATCTGCGGGGGCTCCCTCAT 262
QY 77 eAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPr 97
DB 263 CGCCCCCTCCCGGTCCTCTCCGTGCTCACTGTTTCATGACGAATGGACCTGGAGCC 322
QY 97 oAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAl 117
DB 323 CGGGCCGAGTGGTGGTACTGCTGGCGTCACTCCCGAGGACGGGCCCTGGAGCGGC 382
QY 117 aHisThrArgAlaValAlaAlaIleValValProAlaAsnTrpSerGlnValGluLeuGl 137
DB 383 GCACACCGCGCAGTGGCGCCCATCGTGGTGGCGGCCCACTACAGCCAGTGGAGCTGGG 442
QY 137 yAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPr 157
DB 443 CCGCCAGCTGGCCCTGCTGGCGCTGGCCCTACCCCGCAGCCTGGCGCCCGCTGGGCC 502
QY 157 oValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGl 177

Db 503 TGTCTGCCTGCCCGCCCTCACCCTTGTGTGACGGCAGCCGCTGTGTGGCCACCGG 562
QY 177 YTrpGlyAspVal-GlnGluAlaAspProLeuProLeuProTrpValLeuGln-GluVal 196
Db 563 CTGGGGAGACNTCCANGAGGAGCAATTTCTGTGCTCTCTCCCTGGTGTCTACAGGAAGTG 622
QY 197 GluLeuArgLeuLeuGlyGlu-AlaThrCysGlnCysLeuTyr 210
Db 623 GAGCTAAGTGTGTGGCGAAGGCACTGCTCAATGTCTCTAC 665
RESULT 8
AAK93456
ID AAK93456 standard; cDNA; 670 BP.
XX AC AAK93456;
XX 06-NOV-2001 (first entry)
XX Human cDNA clone representative sequence, SEQ ID NO: 1916.
DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX Homo sapiens.
XX EPI130094-A2.
XX 05-SEP-2001.
XX 07-JUL-2000; 2000EP-00114089.
XX 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
PI WPI; 2001-524255/58.
XX 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX Example 11; SEQ ID NO 1916; 1380pp + Sequence Listing; English.
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence was used as the representative sequence
CC from a human clone which was used in homology searches to identify the
CC clone. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in CD-ROM format directly from
CC EPO
XX
SQ Sequence 670 BP; 97 A; 245 C; 205 G; 118 T; 0 U; 5 Other;
Alignment Scores:
Pred. No.: 1.94e-39 Length: 670
Score: 861.00 Matches: 167
Percent Similarity: 94.92% Conservatives: 1
Best Local Similarity: 94.35% Mismatches: 5
Query Match: 44.09% Indels: 4
DB: Gaps: 0
US-10-037-417-46 (1-357) x AAK93456 (1-670)
QY 38 CysGlyArgProGluProSerAlaArgIleVal-GlyGlySerAsnAlaGlnProGlyTh 57

DB 143 TCGGGCGCCCTGAGCCCTCGCGCCGATCGTGGGGGGGCTCAACGCGCAGCCGGGCAC 202
QY 57 rTrpProTtpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuI 77
DB 203 CTGGCTTGGCAGTGAGCTGCACCATGAGGTGGCCACATCTGGGGGGGCTCCCTCAT 262
QY 77 eAlaProSerTtpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuDuPr 97
DB 263 CGCCCCCTCTGGGTCTCTCCGCTCTCTCACTGTCTTTCATGACGAATGGGACGCTGGAGCC 322
QY 97 oAlaAlaGluTtpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAl 117
DB 323 CGGGCGGAGTGTGTGCTGCTGGCGGTGCACTCCAGGACGGGCCCTTGAAGCGCGC 382
QY 117 aHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGl 137
DB 383 GCACACCGCGCAGTGGCGCCCATGTGTGTGGCGGCACTACAGCACCAGTGGAGCTGGG 442
QY 137 yAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTtpPr 157
DB 443 CGCGACCTGGCGCTGTGGCGCTTCCCGCCAGCTGGGCGCCCGCGGTGGCC 502
QY 157 oValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTtpAlaThrGl 177
DB 503 TGTGTGCTGCGCGCGCTCACACCGCTTGTGTCAGCGGACCGCTGCTGGGCCACCGG 562
QY 177 yTtpGlyAspVal-GlnGluAlaAspProLeuProLeuProTtpValLeuGln-GluVal 196
DB 563 CTGGGGAGACNTCCANGAGGCAAAATCTCTGCTCTCCCTGGGTGTACAGAGAAAGTG 622
QY 197 GluLeuArgLeuLeuGlyGlu-AlaThrCysGlnCysLeuTyr 210
DB 623 GAGCTAAGTGTGGGCGAAGCACTGCTCAATGTCTCTAC 665
RESULT 9
ID ABK13566 standard; cDNA; 456 BP.
AC XX
AC ABK13566;
XX DT 08-MAY-2002 (first entry)
XX DE Human prostatic-like serine protease cDNA #2.
KW Human; prostatic-like serine protease; cytostatic; antiatherosclerotic;
KW virucide; osteopathic; antiinflammatory; vasotropic; neuroprotective;
KW trypsin-like; metastasis; autoimmune lesion; atherosclerosis; gene;
KW renal failure; malignant cell; tumour angiogenesis; enzyme; inflammation;
KW chronic obstructive pulmonary disease; COPD; restenosis; atherosclerosis;
KW neurodegenerative disease; prion protein; infection; amyloid plaque;
KW Genstmann-Straussler Syndrome; viral infection; Scrapie;
KW Creutzfeldt-Jakob disease; metastatic cancer; lipid accumulation;
KW osteoporosis; Paget's disease; ss; EST; expressed sequence tag.
OS Homo sapiens.
XX PN W0200198467-A2.
XX PD 27-DEC-2001.
XX PF 22-JUN-2001; 2001W0-EF007117.
XX PR 23-JUN-2000; 2000US-0213598P.
XX PR 20-MAR-2001; 2001US-0276909P.
XX PA (FARB) BAYER AG.
XX PI Xiao Y, Morozov V;
XX DR WPI; 2002-114576/15.
XX PT Novel human prostatic-like serine protease polypeptide and polynucleotide

PT which can be regulated for treating metastasis of malignant cells,
PT inflammation, atherosclerosis, neurodegenerative disease and infections.
XX Disclosure; Fig 3; 111pp; English.
PS This invention comprises the cDNA and protein sequences of an isolated
XX prostatic-like serine protease and reagents and methods for regulating
CC the human prostatic-like enzyme activity. Prostatic is a trypsin-like
CC serine protease purified from human seminal fluid. An antibody specific
CC for prostatic-like serine protease is useful for immunodetection and
CC diagnosis of micro-metastases, autoimmune lesions and renal failure in
CC biopsy specimens, plasma samples and body fluids. The antibody may be
CC used to modulate enzyme activity in a disease, such as metastasis of
CC malignant cells, tumour angiogenesis, inflammation, chronic obstructive
CC pulmonary disease (COPD), atherosclerosis, neurodegenerative disease
CC (e.g. prion protein amyloid plaques of Genstmann-Straussler Syndrome,
CC Creutzfeldt-Jakob disease, Scrapie) or infection, particularly viral
CC infection. The human prostatic-like serine protease gene provides a
CC therapeutic target of decreasing human prostatic-like serine protease
CC activity, in particular for treating or preventing metastatic cancer. The
CC agonists and antagonists of the nucleotide sequence may be used to mimic,
CC augment and inhibit the enzyme activity which may be useful to treat
CC osteoporosis, Paget's disease and degradation of bone implants
CC particularly dental implants. Altered levels of human prostatic-like
CC serine protease activity inhibits both smooth muscle cell proliferation
CC and lipid accumulation and inhibit the progression of restenosis and
CC atherosclerosis. The nucleic acid sequence is also useful in diagnostic
CC assays for detecting diseases and abnormalities or susceptibility to
CC diseases related to the presence of mutations in nucleic acid sequences
CC which encode the enzyme. The present sequence represents the human
CC prostatic-like serine protease #2 nucleotide sequence of the invention
XX
SQ Sequence 456 BP; 74 A; 151 C; 167 G; 62 T; 0 U; 2 Other;
Alignment Scores:
Fred. No.: 3.56e-36 Length: 456
Score: 799.00 Matches: 147
Percent Similarity: 97.37% Conservativeness: 1
Best Local Similarity: 96.71% Mismatches: 4
Query Match: 40.91% Indels: 0
DB: Gaps: 0
US-10-037-417-46 (1-357) x ABK13566 (1-456)
QY 97 ProAlaAlaGluTtpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGly 116
DB 456 CCGCGCGCGAGTGTGTGCTGCTGCGCGTGTGCTCAAGGACGGGCCCTTGGACGGC 397
QY 117 AlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeu 136
DB 396 GCGCACACCGCGCAGTGGCGCCATCGTGTGGCGCCCACTACAGCCAGTGGAGCTG 337
QY 137 GlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTtp 156
DB 336 GCGCGCGAGCTGGCCCTGCTGGCGCTGGCCCTCACCGCGCAGCTGGCGCCCGCTGGTGG 277
QY 157 ProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTtpAlaThr 176
DB 276 CTGTGTGCTGCTGGCGCGCTTTCAGCGGCTTTCAGCGGACCGCTGGCGCCACC 217
QY 177 GlyTtpGlyAspValGlnGluAlaAspProLeuProTtpValLeuGlnGluVal 196
DB 216 GCGTGGGAGACGTCCAGGAGGACATCTCTGCTCTCTCCCTTGGTGTCTACAGGAAGTG 157
QY 197 GluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPhe 216
DB 156 GAGCTAAGGCTGCTGGCGAGGCCACCTGTCAATGTCTTCTACAGCCAGCCCGCTCCCTTC 97
QY 217 AsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArg 236
DB 96 AACCTCACTCTCCAGATATTGCCAGGAGTGTGTGTGGCTTACCCAGGCGCCGAGG 37
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeuVal 248

US-10-037-417-46 (1-357) x AAS69040 (1-597)

QY 93 GlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyVal-HisSerGlnAspG1 112
Db 3 GGGACGCTGGAGCCGCGCGGCGAGTGGTGGTACTGCTGGGGCGTCAACTCCAGGACGG 62
QY 112 yProLeuAspGlyAlaHisThrArgAlaValAlaAlaLeuValProAlaAsnTyrSe 132
Db 63 GCGCCCTGGACGCGCGCACACCGCGCAGTGGCGCCATCGTGGTGGCGGCGCACTACAG 122
QY 132 xGlaValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuG1 152
Db 123 CCAAGTGAGCTGGCGCGCGAGCCTGGCCCTGCTGGCCCTGAGCTCACCGCGCAGCTGG 182
QY 152 yProAlaValTrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAl 172
Db 183 CCGCGCGCTGGTGGCTGCTGCTGCCCGCGCTCACACCGCTTCGTGTCACGCGACCGC 242
QY 172 aCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTIPVa 192
Db 243 CTGCTGGCGCACCGCTGGGAGAGCTCCAGAGGAGATCTCTGCCCTCTCCCTTGGGT 302
QY 192 lLeuGlnGluValGluLeuArgLeuGlyGluAlaThrCysGlnCysLeuTyrSerG1 212
Db 303 GCTACAGGAAGTGGAGCTAAGCTGCTGGCGAGGCGCACCTGTCAATGTCTTACAGCCA 362
QY 212 nProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrPr 232
Db 363 GCGCGGTCCCTTCAACCTCACTCTCCAGATATTCAGGATGTCGTCGTCGGCTACCC 422
QY 232 oGluGlyArgArgAspThrCysGln---GlyAspSerGlyGlyProLeuValCysGluG1 251
Db 423 AGAGGCGCGCAGGACACCTGCCAGCCAGTCCAGAGCCCGGTCCATGTGCAATCAGCGA 482
QY 251 u 251
Db 483 A 483

RESULT 11
ABT31936
ID ABT31936 standard; DNA; 1733 BP.
XX AC ABT31936;
XX DT 01-MAY-2003 (first entry)
XX DE Human breast cancer / ovarian cancer related coding sequence #43.
XX KW Human; gene; ds; cytostatic; breast cancer; ovarian cancer.
XX OS Homo sapiens.
XX FN WO2003000012-A2.
XX PD 03-JAN-2003.
XX PF 21-JUN-2002; 2002WO-US019773.
XX PR 21-JUN-2001; 2001US-0300159P.
XX PR 27-JUN-2001; 2001US-0301351P.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PI Veiby OP;
XX DR WPI; 2003-267848/26.
XX PT P-PSDB; ABU37067.
XX PT Determining the presence of breast cancer in an individual, involves
XX using specific polynucleotide markers.
XX PS Disclosure; Page 192-193; 233pp; English.
XX

Db 36 GACACCTGCGAGGTGACTCTGGGGGGCCCCCTGGTIC 1

RESULT 10
AAS69040
ID AAS69040 standard; cDNA; 597 BP.
XX AC AAS69040;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #4844.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG04853.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 1; SEQ ID NO 4844; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX CC coding sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 597 BP; 97 A; 217 C; 196 G; 87 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2e-35 Length: 597
Score: 787.50 Matches: 150
Percent Similarity: 94.41% Conservative: 2
Best Local Similarity: 93.17% Mismatches: 7
Query Match: 40.32% Indels: 2
DB: 5 Gaps: 1

particular use with patients having an enhanced risk of developing ovarian cancer (e.g. patients having a familial history of ovarian cancer). The cancer markers may be used in the management and treatment of e.g. brain and central nervous system disorders (e.g. bacterial and viral meningitis, Alzheimer's disease or Parkinson's disease), brain disorders (e.g. cerebral oedema, hydrocephalus or brain herniations), inflammations (e.g. bacterial or viral meningitis or encephalitis), testicular disorders (e.g. nontuberculous granulomatous orchitis), connective tissue disorders, or heart disorders (e.g. ischaemic heart disease or atherosclerosis). The composition and methods may also be used in assessing the histological type of neoplasm associated with ovarian cancer, monitoring the progression of ovarian cancer, determining whether ovarian cancer has metastasized or is likely to metastasize, selecting a composition for inhibiting ovarian cancer, assessing the ovarian carcinogenic potential of a compound, or inhibiting ovarian cancer or at risk of developing ovarian cancer. The present nucleic acid sequence encodes one of the ovarian cancer markers described in the invention

Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5.97e-30 Length: 1834
Score: 697.00 Matches: 160
Percent Similarity: 55.29% Conservative: 49
Best Local Similarity: 42.33% Mismatches: 133
Query Match: 35.69% Indels: 36
Gaps: 11

US-10-037-417-46 (1-357) x ABS76501 (1-1834)

Qy 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 229 ATGGCCAGAGGGGGTCTCGGGGCTGGGAGTGGGGCTGTGGCC-----AlaArgGlyPro 276
Qy 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db 277 ATTCTGCTCTATCTTGGATTACTCGGTTCGGGACAGGAGCGGAGGGGAGAGCTCCC 336
Qy 36 ProTyrCysGlyArgProGluProSerAlaArgLeuValGlyGlySerAlaGlnPro 55
Db 337 -----TGCGGT---GTGGCCCCCAAGCAGCATCACAGGTGGCAGCGTGCAGTCGCC 387
Qy 56 GlyThrTrpProTrpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
Db 388 GGTTCAGTGGCCCTGGCAGGTGAGCATCACCTATGAAAGCGGCTCCATGTGTGTGGTCT 447
Qy 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db 448 CTCGTGTCTGAGCAGTGGGTGTGTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 507
Qy 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 508 GAA-----GCCTATGAGTCAAGTGGGGGCCACAGCTAGACTCTCTCTCTCTCT 558
Qy 116 GlyAlaHisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGlu 135
Db 559 GAGCGCAAGGTGAGCAGCTGAGGACATCATCCCCCAGGAGGAGGAGGAGGAGGAG 618
Qy 136 LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 619 TCCAGGGCGCATGTCATCTCTCCACTCCAGCAGACCCATCACCTCTCCCGCTACATC 678
Qy 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 679 CGGCCCATCTGCTCCCTGCAGCAACGCGCTCTCTCCCAACGCGCTCCAGCTGCATGTC 738
Qy 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuTrpValLeuGlnGlu 195
Db 739 ACTGGTGGGGTCATGTGCCCTCTGAGGAGCTCTGAGCGCCCAAGCCACTGACAGCAA 798
Qy 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215

Db 799 CTCGAGGTGCCTCTGATCAGTCGTGAGACGTGTAACTGCTGTACAAATCGACGCCAAG 858
Qy 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
Db 859 CTTGAGGAGCGGCACCTTTCTCCAAAGAGGACATGGTGTGTGTCTATGTGGAGGGGGC 918
Qy 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGlyGlyArgTrp 255
Db 919 AAGAGCGCTTCCAGGGTGACTCTGGGGGCCCACTCTCTGCGCTGTGGAGGGTCTCTGG 978
Qy 256 PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgAsnArgProGlyVal 275
Db 979 TACCTGAGGGGCATCTGAGCTGGGAGATGCTGTGGGGCCGCAACAGGCCCTGGTGTG 1038
Qy 276 PheThrAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluPro 295
Db 1039 TACACTCTGGCCCTCCAGCATGCTCTCTGATCAAAGCAAGGTG-----ACAGAACTC 1092
Qy 296 GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
Db 1093 CAGCCTCGTGTGTGGTCCCAACCCAGGAGTCCAGCCCGACAGCAACCTCTGTGGCAGC 1152
Qy 313 GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db 1153 CACCTGGCTTTCAGCTCTGCCCCAGCCAGGGCTTGTGAGGCCCCATCTCTTCTGCTCT 1212
Qy 332 ValGlyValSerThrGlyThrLysSerLeuValLeuProTrpLeuSer----- 347
Db 1213 CTGGGCTGTGCTCTGGGC-----CTCCTCTCCCATGCTGCTCAGCGAGCACTGAGCT 1263
Qy 348 -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db 1264 GGCCCTACTTCCAGGTGGATGCATCATCTCAAGGACAGGAGCTGTCTCTTC 1317
RESULT 14
ID ABL67949 standard; DNA; 1834 BP.
XX ABL67949;
XX ABL67949;
DT 15-MAY-2002 (first entry)
DE Ovary cancer related gene sequence SEQ ID NO:6286.
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytosstatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX Homo sapiens.
XX WO200194629-A2.
PN 13-DEC-2001.
XX 30-MAY-2001; 2001WO-US010838.
XX 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.

PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX
 PA (AVAL-) AVALON PHARM.

XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX
 XX WPI; 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 in expression of a gene of a signature gene set.

XX Claim 1; SEQ ID NO 6286; 44pp; English.

XX The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms
 CC tumour

XX Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5,97e-30 Length: 1834
 Score: 697.00 Matches: 160
 Percent Similarity: 55.29% Conservative: 49
 Best Local Similarity: 42.33% Mismatches: 133
 Query Match: 35.69% Indels: 36
 DB: 6 Gaps: 11

US-10-037-417-46 (1-357) x ABL67949 (1-1834)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
 DB ATGCCCAAGAGGGGGTCTCTGGGCGCTGGGAGCTGGGGGCTGTGGCC----- 276

QY 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
 DB ATTCGTCTCTATCTGGATTACTCCGGTCGGGACAGGAGCGGAGGCGAAGCTCCC 336
 QY 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
 DB -----TGCGET---GTGGCCCCCAACAGCAGCATCACAGGTGGCAGCAGTCAGTCGCC 387
 QY 56 GlyThrTrpProTrpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
 DB GGTCACTGGCCCTGGCAGTCCAGCATCACCTATGAAGCGCTCCATGTGTGGTGGCTCT 447
 QY 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
 DB CTTGGTCTGAGCAGTGGGTGCTGTCTCAGCTGTCTCCTCTCCCGAGGAGCACCACAG 507
 QY 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
 DB -----GCCATGAGGTCAAGCTGGGGGCCACACAGCTAGACTCTCTACTCCGAG 558
 QY 116 GlyAlaHisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGlu 135
 DB GAGGCCAAGGTCAGCACCTCGAGGACATCATCCCCCAGCCAGCTACCTCCAGCAGGGC 618
 QY 136 LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
 DB TCCACAGGGCGCATTTGCATCTCTCCAACTCAGCAGACCCCATCACCTTCCCGCTACATC 678
 QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
 DB CGGCCCCATCTGCTCTCCCTGCAGCCCAAGCCCTCTTCCCAACAGCCCTCCACTGCATGTC 738
 QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
 DB ACTGGCTGGGGTCATGTGGCCCCCTCAGTGAGCTCTCTGACGCCCAAGCCACTGCAGCAA 798
 QY 196 ValGluLeuArgLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
 DB CTGAGGTCCCTCTGATCAGCTCGTGAGCGTGAATCTGCTGTACAACTCGACGCCAAG 858
 QY 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
 DB CCTGAGGAGCCGCACTTTGTCGAAGGAGCATGGTGTGTGGTATGTGGAGGGGGGC 918
 QY 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyArgTrp 255
 DB AAGGACCGCTGCCAGGGTGACTCTGGGGGCCCACTCTCTCTGCTGTGGAGGGTCTCTGG 978
 QY 256 PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
 DB TACCTGACGGGCAATTTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGGCTGTGTGTG 1038
 QY 276 PheThrAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluPro 295
 DB TACACTGGCTCCAGCTATGCTCTCTGATCCAAAGCAAGGTG-----ACAGAACTC 1092
 QY 296 GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
 DB CAGCTCGTGTGGTGGCCCCAAACCCAGGATCCCGAGCCCGACAGCAACCTCTGTGGCAGC 1152
 QY 313 GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
 DB CACCTGGCTTCAGCTCTGCCCGCCAGGCGTGTGTGGGGCCCATCTTTTCTCTGCT 1212
 QY 332 ValGlyValSerThrGlyThrLysSerLeuLeuProTrpLeuSer----- 347
 DB CTGGGCTGGCTCTGGGC-----CTCCTCTCCCACTGGCTCAGCGAGCACTGAGCT 1263
 QY 348 -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
 DB GGCCCTACTTCCAGGATGATGATCATCACTCAAGGACAGGAGCCTGCTCTTC 1317

RESULT 15
ABL68512
ID ABL68512 standard; DNA; 1834 BP.

XX AC ABL68512;
XX DT 15-MAY-2002 (first entry)
XX DE Kidney cancer related gene sequence SEQ ID NO:6849.
XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
XX KW gene; ds.
XX OS Homo sapiens.
XX PN WO2001:94629-A2.
XX PD 13-DEC-2001.
XX PF 30-MAY-2001; 2001WO-US010838.
XX PR 05-JUN-2000; 2000US-0209473P.
XX PR 05-JUN-2000; 2000US-0209531P.
XX PR 18-SEP-2000; 2000US-0233133P.
XX PR 18-SEP-2000; 2000US-0233617P.
XX PR 20-SEP-2000; 2000US-0234009P.
XX PR 20-SEP-2000; 2000US-0234034P.
XX PR 20-SEP-2000; 2000US-0234052P.
XX PR 22-SEP-2000; 2000US-0234509P.
XX PR 22-SEP-2000; 2000US-0234567P.
XX PR 25-SEP-2000; 2000US-0234923P.
XX PR 25-SEP-2000; 2000US-0234924P.
XX PR 25-SEP-2000; 2000US-0235077P.
XX PR 25-SEP-2000; 2000US-0235082P.
XX PR 25-SEP-2000; 2000US-0235134P.
XX PR 25-SEP-2000; 2000US-0235280P.
XX PR 26-SEP-2000; 2000US-0235637P.
XX PR 26-SEP-2000; 2000US-0235638P.
XX PR 27-SEP-2000; 2000US-0235711P.
XX PR 27-SEP-2000; 2000US-0235720P.
XX PR 27-SEP-2000; 2000US-0235840P.
XX PR 27-SEP-2000; 2000US-0235863P.
XX PR 28-SEP-2000; 2000US-0236028P.
XX PR 28-SEP-2000; 2000US-0236032P.
XX PR 28-SEP-2000; 2000US-0236033P.
XX PR 28-SEP-2000; 2000US-0236034P.
XX PR 28-SEP-2000; 2000US-0236109P.
XX PR 28-SEP-2000; 2000US-0236111P.
XX PR 29-SEP-2000; 2000US-0236842P.
XX PR 29-SEP-2000; 2000US-0236891P.
XX PR 02-OCT-2000; 2000US-0237172P.
XX PR 02-OCT-2000; 2000US-0237173P.
XX PR 02-OCT-2000; 2000US-0237278P.
XX PR 02-OCT-2000; 2000US-0237294P.
XX PR 02-OCT-2000; 2000US-0237295P.
XX PR 02-OCT-2000; 2000US-0237316P.
XX PR 03-OCT-2000; 2000US-0237425P.
XX PR 03-OCT-2000; 2000US-0237598P.
XX PR 03-OCT-2000; 2000US-0237604P.
XX PR 03-OCT-2000; 2000US-0237606P.
XX PR 03-OCT-2000; 2000US-0237608P.
XX PR 01-NOV-2000; 2000US-0244867P.
XX PR 01-NOV-2000; 2000US-0245084P.
XX PA (AVAL-) AVALON PHARM.
XX PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrikan S;
XX PI Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.

PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.

XX Claim 1; SEQ ID NO 6849; 44pp; English.

XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing calls to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 847 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour

XX Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5, 97e-30 Length: 1834
Score: 697.00 Matches: 160
Percent Similarity: 55.29% Conservative: 49
Best Local Similarity: 42.33% Mismatches: 133
Query Match: 35.69% Indels: 36
Gaps: 11

US-10-037-417-46 (1-357) x ABL68512 (1-1834)

Qy 1 MetAlaGlnIysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 229 ATGGCCCAAGAGGGGGTCTGGGGCTGGGCGAGTGGGGCTGTGGCC----- 276
Qy 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db 277 ATTCTGCTCTATCTTGGATTACTCCGGTCGGGGACAGAGCGGAGGGGCGAGAGCTCCC 336
Qy 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
Db 337 -----TGGGGT---GTGGCCCCCAAGCAGCATCACAGGTGGCAGCAGTCCGCC 387
Qy 56 GlyThrTrpProTrpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
Db 388 GGTGAGTGGCCCTGGCAGGTACGATCACCTATGAAGCGCTCCATGTGTGGTGGCTCT 447
Qy 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db 448 CTCGTGTCTGAGCAGTGGGTGCTGTCTAGCTGCTCAGTCTCTCCCGAGGAGCACCACAAG 507
Qy 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 508 GAA-----GCCTATGAGGTCAAGCTGGGGGCCACCCAGCTAGACTCTCTACTCCGAG 558
Qy 116 GlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGlu 135
Db 559 GACGCCAAGGTCAGCACCTCCGAAGGACATCATCCCCCAGCTACCTCCAGGAGGGC 618
Qy 136 LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 619 TCCAGGGCGGACATTGCACTCTCCACTAGCAGACCCCATCCTTCTCCCGGTACATC 678
Qy 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 679 CGGCCCATCTGCTCCCTGCAGCAACGCTCTTCCCAACGGGCTCCACTGCACACTGTC 738
Qy 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 15:57:04 ; Search time 2489 Seconds
(without alignments)
4283.165 Million cell updates/sec

Title: US-10-037-417-46
Perfect score: 1953
Sequence: 1 MAQKGVLPQLGAVANS...TKSLVLPWLSPHSLGLWGWF 357

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlip
-Q=/cg221/USPTO_spool_p/US10037417/runat_24022004_101038_17460/app_query.fasta_1.519
-DB=EST -QPMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417_CGN 1 1 3549 @runat_24022004_101038_17460 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_hic.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_hic.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: em_gss_hum.*
- 18: em_gss_inv.*
- 19: em_gss_pln.*
- 20: em_gss_vrt.*
- 21: em_gss_fun.*
- 22: em_gss_mam.*
- 23: em_gss_mus.*
- 24: em_gss_pro.*
- 25: em_gss_rod.*
- 26: em_gss_phg.*
- 27: em_gss_vrl.*
- 28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	965	49.4	537	10	AW450407	AW450407 UI-H-BI3-
C 2	799	40.9	456	9	AI190509	AI190509 qq49f07.x
C 3	613	31.6	1699	11	AK078696	AK078696 Mus muscu
C 4	598.5	30.6	1629	11	AK010640	AK010640 Mus muscu
C 5	589	30.2	1071	9	AL551470	AL551470 AL551470
C 6	586.5	30.0	911	13	BX844660	BX844660 BX844660
C 7	573.5	29.4	1150	11	AK080281	AK080281 Mus muscu
C 8	565	28.9	1044	14	CD510038	CD510038 CDA99-F11
C 9	559.5	28.6	819	14	CA588481	CA588481 hab64c07.
C 10	555.5	28.4	1188	11	BC034294	BC034294 Homo sapi
C 11	551	28.2	938	14	CF238383	CF238383 AGENCOURT
C 12	550	28.2	777	13	BX860661	BX860661 BX860661
C 13	540.5	27.7	3035	11	AK004939	AK004939 Mus muscu
C 14	540	27.6	2487	29	AY419856	AY419856 Homo sapi
C 15	533	27.3	905	13	BQ729181	BQ729181 AGENCOURT
C 16	531	27.2	1050	11	AK006271	AK006271 Mus muscu
C 17	530.5	27.2	1323	11	AK014645	AK014645 Mus muscu
C 18	530	27.1	786	13	BX644781	BX644781 DKFZ5781G
C 19	529.5	27.1	946	9	AL578261	AL578261 AL578261
C 20	528.5	27.1	943	13	BQ222964	BQ222964 AGENCOURT
C 21	527.5	27.0	925	13	BX777626	BX777626 BX777626
C 22	523.5	26.8	794	14	CF151097	CF151097 AGENCOURT
C 23	523	26.8	2348	29	AY419858	AY419858 Mus muscu
C 24	523	26.8	3248	11	BC008514	BC008514 Mus muscu
C 25	523	26.8	4017	11	AK052738	AK052738 Mus muscu
C 26	522.5	26.8	908	13	BQ946413	BQ946413 AGENCOURT
C 27	521	26.7	1010	9	BY710051	BY710051 BY710051
C 28	517	26.5	947	9	AL555870	AL555870 AL555870
C 29	517	26.5	1013	12	BM923713	BM923713 AGENCOURT
C 30	516	26.4	784	14	CK029454	CK029454 AGENCOURT
C 31	516	26.4	893	14	CF239681	CF239681 AGENCOURT
C 32	514.5	26.3	969	14	CA488785	CA488785 AGENCOURT
C 33	514	26.3	765	14	CA384091	CA384091 664487 NC
C 34	512	26.2	889	12	BI218460	BI218460 602937954
C 35	510.5	26.1	2335	11	AK045663	AK045663 Mus muscu
C 36	509.5	26.1	861	12	BI763558	BI763558 603050292
C 37	506	25.9	744	14	CF998562	CF998562 AGENCOURT
C 38	506	25.9	1201	13	BX417595	BX417595 BX417595
C 39	505.5	25.9	802	12	BG962187	BG962187 602826954
C 40	505	25.9	881	14	CF242120	CF242120 AGENCOURT
C 41	504.5	25.8	915	13	BUS24192	BUS24192 AGENCOURT
C 42	502.5	25.7	773	14	CK000011	CK000011 AGENCOURT
C 43	498.5	25.5	728	12	BG969117	BG969117 602835094
C 44	498.5	25.5	1172	14	CD508404	CD508404 CDA90-F11
C 45	495	25.3	740	10	AW202972	AW202972 fJ24904.Y

ALIGNMENTS

RESULT 1
AW450407/c
LOCUS
DEFINITION
IMAGE: 2735037 3', mRNA sequence.
ACCESSION
AW450407
VERSION
AW450407.1 GI:6991183
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 537)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 Oligo-dt track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA library Preparation: M.B.
 Soares Lab Clone distribution: NCI-CGAP clone distribution
 Information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbr/image/image.html
 Seq primer: M13 Forward
 POLYA=No.

FEATURES Location/Qualifiers
 source
 1..537
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2735037"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP Sub5"
 /note="vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not 1; Site 2: Eco RI; NCI CGAP Sub5
 is a subtracted library derived from NCI CGAP Sub4. The
 NCI CGAP Sub5 library had 3 million recombinants. A
 single-stranded DNA preparation of NCI CGAP Sub4 was used
 as a tracer in a subtractive hybridization with a driver
 comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM
 3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
 1323376-1323911, 1456008-1456775, 1500552-1502855);
 NCI CGAP Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
 (IMAGE Clones) 1323912-1325831, 1471368-1472903,
 1492104-1493255); NCI CGAP Lu5 pool 1 LLAM 3575-3582,
 3851-3854 (IMAGE Clones)
 1414920-1417991, 1520904-1522439); NCI CGAP GC4 pool 1 LLAM
 3164-3167, 3716-3720, 3733-3735 (IMAGE Clones)
 1257096-1258631, 1469064-1470983, 1475592-1476743);
 NCI CGAP Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
 (IMAGE Clones) 985608-986759, 1101192-1101959,
 1217928-1220615); NCI CGAP_Col0 pool 1 LLAM 2644-2653,
 2871-2872 (IMAGE Clones)
 1057416-1061255, 1144584-1145351). (10% of the driver
 population), plus a pool of 3,840 arrayed clones from
 NCI CGAP Sub1 (IMAGE Clones) 2708616-2710535) and
 NCI CGAP Sub2 (IMAGE Clones) 2710536-2712455) (10% of
 the driver population), plus a pool of 11,136 clones from
 NCI CGAP Sub3 (IMAGE Clones) 2712456-2723591) (10% of the
 driver population), plus a pool of 5,472 clones from
 NCI CGAP Sub4 (IMAGE Clones) 2723592-2728969) (70% of the
 driver population). Subtraction was performed as
 previously described (Bonaldi, Lennon & Soares (1996):
 Normalization and Subtraction: Two Approaches To
 Facilitate Gene Discovery. Genome Research 6, 791-806.
 TAG_TISSUE=colon
 TAG_LIB=NCI CGAP_Col0
 TAG_SEQ=AAACG"

ORIGIN
 Alignment Scores: 1.46e-66 Length: 537
 Pred. No.: 965.00 Matches: 177
 Score: 99.44% Conservative: 0
 Percent Similarity: 99.44% Mismatches: 1
 Best Local Similarity: 99.44% Indels: 0
 Query Match: 49.41% Gaps: 0
 DB: 10

US-10-037-417-46 (1-357) x AW450407 (1-537)

Qy 76 Leu1eAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
 Db 536 CTCATCGCCCCCTCTGGTCTCTCCGCTCTACTGTTTCATGCGAATGGGACGTG 477
 Qy 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115

Db 476 GAGCCGCGCCGAGTGGTTCGTACTGCTGGCGTGCATCTCCAGACGCGGCCCTGGAC 417
 Qy 116 GlyAlaHisThrArgAlaValAlaAlaValValProAlaAsnTyrSerGlnValGlu 135
 Db 416 GCGCGGCACACCCGCGAGTCGCGCCATCGTGGTGCCGCCCACTACAGCCAACTGGAG 357
 Qy 136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
 Db 356 CTGGGCGCGACCTGGCCCTGCTGGCCTGGCTCACCCGCGCAGCTGGGCCCGCCGTG 297
 Qy 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
 Db 296 TGCCCTGTCTGCTGCGCCGCGCTCACACCGCTTGTGTCACGCGCACCCGCTGCTGGGCC 237
 Qy 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
 Db 236 ACCGGCTGGGGAGAGCTCCAGGAGGAGATCTCTCCCTCTCCCTGGGTGCTACAGGA 177
 Qy 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
 Db 176 GTGGACCTAAGGCTGCTGGCGAGGCCACTGTGTCAATGTCTCTACAGCCAGCCGCTCCC 117
 Qy 216 PheAsnLeuThrLeuGlnLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
 Db 116 TTCAACCTCCTCTCCAGATATTGCCAGGATGCTGTGTGCTGGCTACCCAGAGGGCCGC 57
 Qy 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGlyGly 253
 Db 56 AGGACACCTGCCAGGTGACTCTGGGGGCGCCCTGCTGTGAGGAGGCGGC 3

RESULT 2
 A1190509/c
 LOCUS A1190509
 DEFINITION qd49f07.x1 Soares fetal heart NbH19W Homo sapiens cDNA clone
 IMAGE:1732837 3' similar to SW:MCPE_MOUSE P21845 MAST CELL PROTEASE
 6 PRECURSOR ;, mRNA sequence.
 ACCESSION A1190509
 VERSION A1190509.1 GI:3741718
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 456)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 712 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 452.
 Location/Qualifiers
 1..456
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1732837"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares fetal heart NbH19W"
 /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
 modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAAGTGGGCGCGCATCTTTTTTTTTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBHL19W."

ORIGIN

Alignment Scores:
Pred. No.: 1.65e-53 Length: 456
Score: 799.00 Matches: 147
Percent Similarity: 97.37% Conservative: 1
Best Local Similarity: 96.71% Mismatches: 4
Query Match: 40.91% Indels: 0
DB: 9 Gaps: 0

US-10-037-417-46 (1-357) x A1190509 (1-456)

Qy 97 ProAlaAlaGluTyrSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGly 116
Db 456 CCGCGCGCGAGTGGTGGTACCTGGTGGCGGTGACCTCCAGGACGGGCCCTGGACGGC 397
Qy 117 AlaHisThrArgAlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeu 136
Db 396 GCGCACCGCGAGTGGCGCCATCGTGGTGGCGGCACTACAGCCAGTGGAGCTG 337
Qy 137 GlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyr 156
Db 336 GCGCGCGACCTGGCGCTGCTGGCGTGGCTCACCAGCGAGCTGGCGCGCGCTGGTGG 277
Qy 157 ProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTyrAlaThr 176
Db 276 CCTGTGCTGCTGGCGCGCTCACCAGCGTGGTGGCGAGCTGGTGGCGCGAGCTGG 217
Qy 177 GlyTyrGlyAspValGlnGluAlaAspProLeuProTyrValLeuGlnGluVal 196
Db 216 GGCTGGGAGACGCTCCAGGAGGAGATCTGTGCTCTCCCTGGTGGTGTACAGAGTGT 157
Qy 197 GluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPhe 216
Db 156 GAGCTAAGGCTGCTGGCGAGCGCCACCTGTCAATGTCTCTACAGCCAGCGCGTCTTC 97
Qy 217 AsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 236
Db 96 AACCTCACTCTCCAGATATCCAGGAGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 37
Qy 237 AspThrCysGlnGlyAspSerGlyGlyProLeuVal 248
Db 36 GACACCTGCCAGGTGACTCTGGGGGGCCCTGTCTC 1

RESULT 3
AK078696
LOCUS
DEFINITION AK078696 1699 bp mRNA linear HTC 18-SEP-2003
library, clone:7530415E19 product:PROSTASIN, full insert sequence.
ACCESSION AK078696
VERSION AK078696.1 GI:26347426
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE
PUBMED
REFERENCE
AUTHORS

20499374
11042159
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kiteunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishire,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1699)
AUTHORS
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

MEDLINE
PUBMED
REFERENCE
AUTHORS

20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1699)
AUTHORS
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

MEDLINE
PUBMED
REFERENCE
AUTHORS

20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1699)
AUTHORS
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

MEDLINE
PUBMED
REFERENCE
AUTHORS

20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1699)
AUTHORS
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

COMMENT

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Please visit our web site for further details.

URL:http://genome.gsc.riken.go.jp/

URL:http://fantom.gsc.riken.go.jp/

Location/Qualifiers

1. 1699
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:7530415E19"
/db_xref="MGI:2396949"
/db_xref="taxon:10090"
/clone="7530415E19"
/sex="male"
/tissue_type="eyeball"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
184. 1206
/notes="unnamed protein product; PROSTASIN (SPTR|BAB82496,
evidence: FASTY, 98.5%ID, 100%length, match=1020)
putative"
/codon_start=1

FEATURES
source

CDS

```

/protein_id="BAC37362.1"
/db_xref="GI:26347427"
/translation="MALRVGLGLQLQLEAVTILLGLGOSIRADGTASCGAVIQPR
ITGGQWQSPGQWPQWQVITYDGNHYCVGSLVSNKVVSAHCFPREHSREVEVKLG
DHQDSDYSNDTVVTHVAQIILTHSYREBSQGIALIRLSSPVTFSRIRPICLPPAN
ASPFNGHCVTGWGHVAPSVSQTDRPQQLQLEVPILSRITSCSLININAVPEEPTI
QDMLCAGYVKGQDAQCGDSGGFLSCPMEGIWYLAGIVSWGDACGAPRPGVYTLTS
TYASWTHHHVHLQPRVVPQTOBSQPDGHLCNEHPVFSSAAAPKLLRPLFLPLGLTL
GLLSLA"
polya_signal 1671..1676
/feature="putative"
polya_site 1699
/feature="putative"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 7,22e-38 Length: 1699
Score: 613.00 Matches: 145
Percent Similarity: 54.73% Conservative: 46
Best Local Similarity: 41.55% Mismatches: 140
Query Match: 31.39% Indels: 18
DB: 11 Gaps: 10

```

US-10-037-417-46 (1-357) x AK078696 (1-1699)

```

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 184 ATGGCCCTAAGGGTGGGCTGGGACTTGGGAGCTGAAGCTGTGACC-----ATT 234
QY 21 TyrSerLeuTyrGlyLeuValProSerGly----ProAlaArgGly----ProProTyrCys 38
Db 235 CTGCTCTCTCGAATGCTCCAGCTGGGAATCCGAGCTGACGGGAGCTGAAGCCCTCTGT 294
QY 39 GlyArgPro---GluProSerAlaArgIleValGlyGlySerAsnAlaGln---ProGly 56
Db 295 GGTGCCCTCATCGACCA-----CGCATACCGGTGTGGCGAGTGGCAAGCCCGGT 348
QY 57 ThrTyrProTyrGlnValSerLeuHisGlyGlyHisLeuGlyHisLeuGlySerLeu 76
Db 349 CAGTGGCCCTGGCAGGTGAGCTACCTACGATGCGCAACCATGTTTGTGGCGGTGCTC 408
QY 77 IleAlaProSerTyrValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGlu 96
Db 409 GTGTCAATAAATAGGTGGTGTCTGCTGCTCAGTCTCCCGACAGAACACAGCAGGAA 469
QY 97 ProAlaAlaGluTyrSerValLeuGlyValHisSerGlnAspGlyProLeuAspGly 116
Db 469 -----GGTATGAGGTGAAGCTGGGGGACCAACAGCTAGACTCTCAGCAATGAC 519
QY 117 AlaHisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGluLeu 136
Db 520 ACTGTGTGCCACAGTGGCTCAGATCATCATCCCACTCAAGCTACCGAAGAGGGCTCC 579
QY 137 GlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyr 156
Db 580 CAGGGGACATCGCGCTCATCGGCTCAGCAGTCTGTGCTCTCCCGTACATCAGA 639
QY 157 ProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTyrAlaThr 176
Db 640 CCCATCTGCTCCCTCGCAGCAATAGCTCTTCCCAAGGSCCTTCACTGTACTGTCA 699
QY 177 GlyTyrGlyAspValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGluVal 196
Db 700 GGATGGGTGATGTGGCTCTCTTCAAGCCCTCCAGACCCCTAGGCTCTCTCAGCAGCTC 759
QY 197 GluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPhe 216
Db 760 GAGGTACCACATCATCGCGGGAACCTGTAGCTGCTGTACACATTAATGCGGTGCT 819
QY 217 AsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArg 236
Db 820 GAAGAACCGCACATATCCAGCAGGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 879

```

```

QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGlyGlyArgTyrPhe 256
Db 880 GATGCGCTGCCAGGGTGACTCTCGGGGCCCCACCTCTCTTGTCCCATGAGGCGCATCTGGTAC 939
QY 257 GlnAlaGlyIleThrSerPheGlyGlyCysGlyArgArgAsnArgProGlyValPhe 276
Db 940 TTGGCAGGCAATTGTAGTTGGGGTGATGCTGTGGTGGCCCCCAACAGGCTGGAGTATAC 999
QY 277 ThrAlaValAlaThrTyrGluAlaThrIleArgGluGlnValMetGlySerGluProGly 296
Db 1000 ACTCTGACTTCTACCTATGCTTCTGATCCACCACCATGTG-----GCAGAGCTCCAG 1053
QY 297 ProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHisGln 313
Db 1054 CCTCGAGTTGTCCCCAGACTCAGGAGTCCAGCCGATGGCCACCTCTGCAACCATCAT 1113
QY 314 ThrAlaPheLeuAspSerAlaArg---IleLeuLeuArgProLeuSerHisLeuSerVal 332
Db 1114 CCGTCTCTCAGCTCAGCGGAGCCGCCGAAATGTTAAGGCCCGTACTTTCTCGGCACATT 1173
QY 333 GlyValSerThrGlyThrLysSerLeu 341
Db 1174 GGTCTGACCTCGGCTCTGTCTTTG 1200

RESULT 4
AK010640 1629 bp mRNA linear HTC 20-SEP-2003
LOCUS Mus musculus ES cells cDNA, RIKEN full-length enriched library,
DEFINITION clone:2410039E18 product:PROTASIN, full insert sequence.
ACCESSION AK010640.1 GI:12846228
VERSION AK010640.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
3
REFERENCE
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Iehi,Y., Nakamura,S., Hazama,M., Nishine,T., Hatada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwake,S., Inoue,K., Iizawa,Y., Tanaka,T., Tanaka,T., Matsura,S., Kawai,J.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
4
REFERENCE
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED 11076861
5
REFERENCE
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

```

JOURNAL
REFERENCE
AUTHORS

Nature 420, 563-573 (2002)
6 (bases 1 to 1629)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

TITLE
JOURNAL

COMMENT

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5',
GAGAGAGATCTCGATTCTGAATTAATTAATCCCCCCCCC 3']. cDNA was cleaved
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.

FEATURES
source

Location/Qualifiers
1..1629
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:2410039E18"
/db_xref="WGI:1910128"
/db_xref="taxon:10090"
/clone="2410039E18"
/cell_type="ES cells"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
115..892
/note="unnamed protein product; PROSTASIN (SPTR)[BAB82496,
evidence: FASTY, 98.5%ID, 100%length, match=1020]
putative"

CDS

/codon_start=1
/protein_id="BAC25310.1"
/db_xref="GI:26370154"
/translation="MALRVGLGLQLEAVTILLIGLQSRADGTEASCGAVIQPR
ITGGSAKPGQWPQVSTYDGNHVCSSIVSNKVVSAACFPREHREAYEVKLGA
HQLSDYNDTVVHTVAQIITHSYREBSQGDIALIRSSPVTSRYRIRPCDPAANA
SFNGLHCTVGTGWGVAPSVLSQTPLOQLEVLPLISRETCSCLYNINAVPEEPTIQ
QDMLCAGYVKGKMPARTVLGAHSLVPRASGTQAL"

ORIGIN

Alignment Scores:

Pred. No.:	9-55e-37	Length:	1629
Score:	598.50	Matches:	147
Percent Similarity:	55.59%	Conservative:	47
Best Local Similarity:	42.12%	Mismatches:	136
Query Match:	30.65%	Indels:	20
DB:	11	Gaps:	11

US-10-037-417-46 (1-357) x AK010640 (1-1629)

Qy 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20

Db	Qy	115	ATGGCCCTAAGGGTGGGCTGGGACTTGGGAGCTTGGAGCTGTGACC-----ATT	165
Qy	21	TyrSerLeuTyrGlyLeuValProSerGly---ProAlaArgGly---ProProTyrCys	38	
Db	166	CTGCTCCTTCGCGAATCTCCAGTCGGGAATCCGAGCTCGCGGACTGAAGCCCTCCTGT	225	
Qy	39	GlyArgPro---GluProSerAlaArgTleValGlyGlySerAsnAlaGlnProGlyThr	57	
Db	256	GGTGGCGTCATCCAGCCA-----CGCATCCCGGTGGTGGCAGTGGCAAGCCCGGTGAG	279	
Qy	58	TppProTppGlnValSerLeuHisGlyGlyHisGlyHisGlyCysGlyGlySerLeuTle	77	
Db	280	TGGCCCTGGCAGGTGAGCATCCTTACGATGGCAACCATGTTTGTGGGGGGTGGTCTG	339	
Qy	78	AlaProSerTppValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro	97	
Db	340	TCAATAAATGGGTGGTGTCTGCTGCTCCTGCTCCAGAGAAACACAGCAGGGAA---	396	
Qy	98	AlaAlaGluTppSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla	117	
Db	397	-----GGTATGAGGTGAAGCTGGGGGCCACCAGCTAGACTCCTACAGCAATGACACT	450	
Qy	118	HisThrArgAlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuGly	137	
Db	451	GTGGTCCACAGTGGGTCTCAGATCATCACCCTCAGCTACCGAAGAGAGGGCTCCAG	510	
Qy	138	AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTppPro	157	
Db	511	GGGACATCGCGCTCATCCGCTCAGCAGTCTGTACCTTCTCCCGGTACATCAGAGCC	570	
Qy	158	ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTppAlaThrGly	177	
Db	571	ATCTGCTCCTCGCAGCAATGCCTCTTCCCAAGCGCTTCTACTGTCTACGGGA	630	
Qy	178	TppGlyAspValGlnGluAlaAspProLeuProLeuProTppValLeuGlnGluValGlu	197	
Db	631	TGGGGTCATGTGGCTCTCTCAGTGAGCTCCAGACCCCTAGGCTCTGTGAGCTGAG	690	
Qy	198	LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn	217	
Db	691	GTACCCTCATCAGCGGGGAACCTGTAGCTGCTTACACATTAATGCGGTGCTGAA	750	
Qy	218	LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp	237	
Db	751	GAACCGCACACTATCCAGCAGGACATGCTGTGTGTGCTATGT- GAAGGAGGCAAGAT	809	
Qy	238	ThrCysGlnGlyAspSerGlyProLeuValCysGluGluGlyGlyArgTppPheGln	257	
Db	810	GCCTGCCAGGTGACTCTGGGGGGCCACTCTCTTGTCCCATGGAGGGCATCTGTACT	869	
Qy	258	AlaGlyIleThrSerPheGlyPheGlyCys---GlyArgArgAsnArgProGlyValPhe	276	
Db	870	GCAGGCATGTGAGTGGGGT---GGATGCTGTGTGCCCCCAACAGGCTGGAGTATAC	926	
Qy	277	ThrAlaValAlaThrTyrGluAlaThrIleArgGluGlnValMetGlySerGluProGly	296	
Db	927	ACTCTGACTTCTACTATGCTTCTGGATCCACCACCATGTG-----GCAGAGCTCCAG	980	
Qy	297	ProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHisGln	313	
Db	981	CCTGAGTTGTCCCCCAGACTCAGGAGTCCCGAGCCGAGTGGCCACTCTGCAACCATCAT	1040	
Qy	314	ThrAlaPheLeuAspSerAlaArg---IleLeuLeuArgProLeuSerHisIleSerVal	332	
Db	1041	CCTGCTTTCAGCTCAGCGGAGCCGCCGAATTTGTTAAGGCCCGTACTTTTCTGCCACTT	1100	
Qy	333	GlyValSerThrGlyThrLysSerLeu 341		
Db	1101	GGTCTGACCTGGGCTCTTGTCTTTG 1127		

RESULT 5
AL551470
LOCUS

1071 bp mRNA linear EST 31-MAY-2003

```

DEFINITION AL551470 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CSODI064YF02 5-PRIME, mRNA sequence.
ACCESSION  AL551470
VERSION     AL551470.2  GI:31273286
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS     Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12889449.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 7896.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CSODI064DC01QP1&cluster=7896.r. Contact :
            Rong Liang Email : liliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CSODI064DC01QP1.
FEATURES    Location/Qualifiers
            source          1..1071
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="CSODI064YF02"
                        /tissue_type="PLACENTA"
                        /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                        /note="First strand cDNA was primed with a NotI-oligo(dT)
                        primer. Five prime end enriched, double-strand cDNA was
                        digested with Not I and cloned into the Not I and EcoR V
                        sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.:      2,81e-36      Length:      1071
Score:          589.00      Matches:      131
Percent Similarity: 57.73%      Conservative: 37
Best Local Similarity: 45.02%      Mismatches:  108
Query Match:    30.16%      Indels:      16
DB:             Gaps:        6

US-10-037-417-46 (1-357) x AL551470 (1-1071)

Qy      1 MetAlaGlnIysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db      148 ATGCCCAAGAGGGGCTCTGGGCGCTGGGAGCTGGGGCTGTGGCC----- 195
Qy      21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db      196 ATTCGTCTATCTATTGTGATTACTCCGTCGGGGACAGAGCGGAGCGGAGCTCCC 255
Qy      36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
Db      256 -----TGGCGT---GTGGCCCCCAGACGACATACAGTGGCAGCATGTGCGTCC 306
Qy      56 GlyThrTrpProTyrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
Db      307 GGTCAATGGCCCTGGCAGTCCAGCATCACCCTATGAAGCGCTCCATGTGTGTGGTCT 366
Qy      76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db      367 CTCGTGTCAGACAGTGGGTCTGTGACGTCTCACTGCTTCCCACGCGACCAACAAG 426
Qy      96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db      427 GAA-----GCCATAGAGTCAAGCTGGGGGCCACAGCTAGACTCTCTCCGAG 477

```

```

Qy      116 GlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGlu 135
Db      478 GAGCCCAAGGTGAGCACCCTGAGGACATCATCCCCACCCAGCTACCTCCAGAGGGC 537
Qy      136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db      538 TCCACAGGGCGACATGTGCTCTCCCAACTCAGCAGACCCATCACTTCTCCCGCTACATC 597
Qy      156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db      598 CGGCCCATCTGCTCTCTCGAGCCCAACGGCTCTCTCCCAACAGCGCTCCACTGCATGTC 657
Qy      176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGlu 195
Db      658 ACTGGCTGGGTGATGTGGC-CCCTCAGTGAGCTCTCTGACGCCCAAGCCACTGCAGCAA 716
Qy      196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db      717 CTCAGGTGCTCTGATCAGTCTGAGAGTGTGTAACATGCTGTACACATCGACGCCAAG 776
Qy      216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
Db      777 CTGAGAGCGCGCACCTTTGTCCAAGAGACATGTGTGTCTGGCTATGTGGAGGGGGC 836
Qy      236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGlyArgTyr 255
Db      837 AAGACCGCTGCCAGGGTGACTCTGGGGGCCACTCTCTCGCTGTGGAGGGTCTCTGG 896
Qy      256 PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db      897 TACCTGACGGGCATTTGTGAGCTGGGAGATGCTGTGGGGCCGCCAACAGCGCTGTGTG 956
Qy      276 PheThrAlaValAlaThrTyrGluAlaTyrIle 286
Db      957 TACACTGCGCTCCAGCTATGCTCTCTGGATC 989

RESULT 6
BX844660          911 bp      mRNA      linear      EST 11-DEC-2003
BX844660 NICHU_XGC_Emb4 Xenopus laevis cDNA clone IMAGp998C059551 ;
IMAGE:4203196 5', mRNA sequence.
ACCESSION  BX844660
VERSION     BX844660.1  GI:39734664
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
ORGANISM    Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
            Xenopodinae; Xenopus.
            1 (bases 1 to 911)
            Heil, O., Neubert, P., Peters, M., Radelof, U., Schneider, D.,
            Schroth, A., Korn, B. and Landgrebe, J.
            Xenopus laevis UniGene Set 1 (RZPDLIB No.988)
            Unpublished (2003)
            Contact: Ina Rolf
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD: IMAGp998C059551.
            RZPDLIB: I.M.A.G.E. cDNA Clone Collection (amp- resistant) (RZPDLIB
            No.988) http://www.rzpd.de/cgi-
            bin/products/showlib.pl.cgi/response?libNo=998 RZPDLIB; Xenopus
            laevis UniGene Set 1 (RZPDLIB No.988) http://www.rzpd.de/cgi-
            bin/products/showlib.pl.cgi/response?libNo=988 Contact: Ina Rolf
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel.: +49 30 32639 101
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            SP6, 5' ATTAGTGACACTATAG 3'.
            Location/Qualifiers
FEATURES

```


source

```

1. .911
/organism="Xenopus laevis"
/mol type="mRNA"
/db xref="taxon-8355"
/clone="IVAG998C059551 ; IMAGE:4203196"
/dev stage="embryo, stage 31-32"
/lab host="DH10B (phage-resistant)"
/clone lib="NICHD XGC Emb4"
/notes="Organ: whole embryo; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dr. Average insert size 2.1 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection
(XGC) library."

```

ORIGIN

Alignment Scores:

```

Pred. No.: 3.44e-36 Length: 911
Score: 586.50 Matches: 118
Percent Similarity: 57.09% Conservatives: 47
Best Local Similarity: 40.83% Mismatches: 97
Query Match: 30.03% Indels: 27
DB: 13 Gaps: 7

```

US-10-037-417-46 (1-357) x BX844660 (1-911)

```

Qy 10 GlyGlnLeuGlyAlaValAlaAsnSerAspSerTyrSerLeuTyrGlyLeuValProSer 29
Db 71 GGGTCGACGAGGAGCATCATCATCAAGTTCTGAAGAACATATGCTCTA-----121
Qy 30 GlyProAlaArgGlyProToTyrCysGlyArgProGluProSerAlaArgIleValGly 49
Db 122 -----TCAGTCGGGTGCGCAATGGTATCAGAACCGGATAGTAGGG 163
Qy 50 GlySerAsnAlaGlnProGlyThrTrpTrpGlnValSerLeuHisGlyGlyGly 69
Db 164 GGAACGACACATGAAGAGGGGGAATGGCCCTGGGACATCAGTCGTGAGCTACAAAGGGGAA 223
Qy 70 HisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPhe 89
Db 224 CCTGTCGCGGGGATCCCTTATTGCCAACAGCTGGATTCTGCTGCTCACTGCTTT 283
Qy 90 MetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSer 109
Db 284 GATTTCG-----CAAAACGTTTCCCAATATAAAGTTTACCTTGGGGTTAT---328
Qy 110 GlnAspGlyProLeuAspGlyAlaHisThr-----ArgAlaValAlaAlaIleValVal 127
Db 329 CGATTGCTCTCTCCCTTCAAAGTCCCAACACACTCTCCAGGAGTGTGAAGAAATCATATA 388
Qy 128 ProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSer 147
Db 389 CACCCAGATTACCAATTGAAGGATCTAATGGGACATTCGACTGATTGAATGGATCAA 448
Qy 148 ProAlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPhe 167
Db 449 CCGTCACCTTTACCCCTTACATTTCTCCAGCCCTCCCTCCACCCCGCTGCTCTGCTT 508
Qy 168 ValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeu 187
Db 509 CTGTCAGGGGTGAATGCTGGTACTGGTGGGAGATATCAAGAAGGACAGCCTCTA 568
Qy 188 ProLeuProTrpValLeuGlnGlnValGluLeuArgLeuGlyGluAlaThrCysGln 207
Db 569 GGTAAACCTTAAACCTTCGAGAAGGCTACAGTGTCTTAATTGCTGCGATAGCTGTGAG 628
Qy 208 CysLeuTyrSer-----GlnProGly-----ProPheAsnLeuThrLeuGln 221
Db 629 TCCATGTACGAAACCCAGCTTGGGATACAAACCTTAATGTTCTCTTT-----673
Qy 222 IleLeuProGlyMetLeuCysAlaGlyTyrProGlnGlyArgArgAspThrCysGlnGly 241
Db 674 ATCTCGTGTACATGTTCTGTGCTGGCTACAGGAAGGAAGATAGACGCTGTGAGGA 733

```

```

Qy 242 AspSerGlyGlyProLeuValCysGluGluGlyArgTrpPheGlnAlaGlyIleThr 261
Db 734 GACTCTGGAGGACCTCTTGTATGTCAGAGTCAACACACATGGTGGCAATACGCGCATTTGTC 793
Qy 262 SerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAlaThr 281
Db 794 AGCTGGGGTATTGATGTGGACAACTAACCAACCGGGTGTCTATACCAAGTCCAGTAT 853
Qy 282 TyrGluAlaTrpIleArgGluGlnVal 290
Db 854 TATGATGCTTGGATCAAGCAGTACATT 880

RESULT 7
AK080281 1150 bp mRNA linear HTC 19-SEP-2003
LOCUS Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
DEFINITION library, clone:A630023F11 product:similar to MARAP SIN PRECURSOR (BC
3.4.21.-) [Homo sapiens], full insert sequence.
ACCESSION AK080281.1 GI:26348448
VERSION HTC; CAP trapper.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kusunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yaneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuura, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnata, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

```

Soqabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yaeunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

TITLE JOURNAL

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

FEATURES source

Location/Qualifiers
1. 1150
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A630023F11"
/db_xref="MG1:2406613"
/db_xref="taxon:10090"
/clone="A630023F11"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
66. 1152

/note="unnamed protein product; putative
similar to MARAPIN PRECURSOR (EC 3.4.21.-) [Homo sapiens]
[SWISSPROT:Q9B0R3, evidence: FASTV, 76.5%ID, 100%length,
match=876]"
/codon_start=1
/protein_id="BAC37864.1"
/db_xref="GI:26348449"
/translation="MRPHIAALILLPLLSRSGEGARTILRACGHPFNRNMYGGENA
LEGHPWQVSIQRNGIHFCGSLIAPTWLTAHCFNSIDISIVQLLGLALQDPG
PHLYVYVKQVKNPQVQWASSADVALVELQGPVTNVLVLCVLPDPVIFESGNV
CWVTGSGSEQDRLNPRVLQKLAVIDIPEKNLLYNKDVESDFOLKTKDMLCA
GFAGKKDACKDGGGLVCLVDQSWQVQAGVISWEGCARNRNRPVTVIRVISHKMT
QIIPELOFQGRAGTQQQKDSQGOORLAGNSAPCLAAHAWLALGALLLRIV"

CDS

/note="unnamed protein product; putative
similar to MARAPIN PRECURSOR (EC 3.4.21.-) [Homo sapiens]
[SWISSPROT:Q9B0R3, evidence: FASTV, 76.5%ID, 100%length,
match=876]"
/codon_start=1
/protein_id="BAC37864.1"
/db_xref="GI:26348449"
/translation="MRPHIAALILLPLLSRSGEGARTILRACGHPFNRNMYGGENA
LEGHPWQVSIQRNGIHFCGSLIAPTWLTAHCFNSIDISIVQLLGLALQDPG
PHLYVYVKQVKNPQVQWASSADVALVELQGPVTNVLVLCVLPDPVIFESGNV
CWVTGSGSEQDRLNPRVLQKLAVIDIPEKNLLYNKDVESDFOLKTKDMLCA
GFAGKKDACKDGGGLVCLVDQSWQVQAGVISWEGCARNRNRPVTVIRVISHKMT
QIIPELOFQGRAGTQQQKDSQGOORLAGNSAPCLAAHAWLALGALLLRIV"

ORIGIN

Alignment Scores:
Pred. No.: 5,32e-35 Length: 1150
Score: 573.50 Matches: 133
Percent Similarity: 52.05% Conservative: 45
Best Local Similarity: 38.89% Mismatches: 133
Query Match: 29.37% Indels: 31
DB: 11 Gaps: 9

US-10-037-417-46 (1-357) x AK080281 (1-1150)

QY 26 LeuValProSerGlyPro-----AlaArgGlyProProTyrCysGlyArgProGluPro 43
Db 108 CTGCTCAGTCTGGAGTCTGAGGGCCAGGAGTCTGAGAGCCTGTGGCCATCCAAAGATG 167
QY 44 SerAlaArgIleValGlySerAsnAlaGlnProGlyThrTrpProTyrGlnValSer 63
Db 168 TTCACCGGATGTTGGCGGGGAGAACGCTTAGAAGGTGAGTGGCCCTGGCAGGTCAAG 227
QY 64 LeuHisHisGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeu 83
Db 228 ATCCAGCGCAGCGGAATCCACTTCTGTGGGGGAGTCTCATCGCACCAACATGGGTCTTC 287

QY 84 SerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerVal 103
Db 288 ACTGCTGCACACTGCTTCCCAACACTTCGGACATA-----TCCATATACCAAGGTC 338
QY 104 LeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAla 123
Db 339 CTGCTGGGGGCACCTAAGCTGCAGCAGCA-----GGACCAACACGCTTGTATGTCC 392
QY 124 AlaIleValValProAlaAsn-----TyrSerGlnValGluLeuGlyAlaAspLeuAla 141
Db 393 GTGAGCAGGTGAAGCAATCTCAGTCAAGGATGGCTCCAGTCCGACGCTGGT 452
QY 142 LeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeuPro 161
Db 453 CTGCTGGAGCTGCAGGGGCGGTGACCTTCAACAATTACATCTCTCTGTGTCTTCT 512
QY 162 ArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspVal 181
Db 513 GATCCCTCGGTCTATCTTGTAGTCGGCATGAATCTGCTGGTCACTGGTGGGACGCC 572
QY 182 GlnGluAlaAspProLeuProTrpValLeuGlnGluValGluLeuArgLeuLeu 201
Db 573 AGTGAACAGACCGACTACCCCAACCCACGGTCTCTGCAGAACTTCTGTGCCCATCA 632
QY 202 GlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeuGln 221
Db 633 GACACGCCCAAGTGCACCTGCTGTACAAAGATGTTGAGTCTGACTTCCAGCTCAA 692
QY 222 ---IleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240
Db 693 ACCATCAAGATGACATGCTGTGCAGGCTTCGCAGAGGGCAAGAGACGCTGTAG 752
QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIle 260
Db 753 GGTGATCTTGGAGGCCCATCTGGTGTCTGTGGACCATGCTGGTGGAGGTGGGTG 812
QY 261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
Db 813 ATCAGCTGGAGGAGGGTGTGTGCCCGTCGGAATCGCCAGGTGTCTACATCCGTGTG 872
QY 281 ThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGlu---ProGlyProAlaPhe 299
Db 873 TCTATCACAATGGAATCCCAAAATCATCCCAAGACTGCAGTTTCAGGGGAGGACG 932
QY 300 ProThrGlnProGlnLys-----ThrGlnSerAsp 309
Db 933 ACCCAGCAGCAACAAAAGACTCCCAAGGTCAGCAGCGCTGGCGGGAATCCGCCCT 992
QY 310 CysLeuHisGlnThrAlaPheLeuAspSerAlaArgIleLeuLeuArgProLeuSerHis 329
Db 993 TGGCTG-----GCTGCCCATGCATGTGTGTGGCTGGCCCTGGGAGCA 1031
QY 330 IleSerValGlyValSerThrGlyThrLysSerLeuValLeuProTrpLeuSerProHis 349
Db 1032 CTGCTGCTCAGATCGTCTAGGGCTGCCAGCC-----CCACAGGTGGAGCCCCAC 1082
QY 350 SerLeu 351
Db 1083 CCACGTG 1088

RESULT 8

CD510038
LOCUS
DEFINITION
CD499-E11.xid-t SHGC-CDA Gasterosteus aculeatus cdna clone
CD499-E11 5', mRNA sequence.
CD510038
VERSION
CD510038.1 GI:31440607
EST.
KEYWORDS
SOURCE
ORGANISM
Gasterosteus aculeatus (three spined stickleback)
Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;

Sequencing by: Washington University Genome Sequencing Center Clone
distribution: Fugu clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 490.

FEATURES

Location/Qualifiers
1. 819
/organism="Takifugu rubripes"
/mol_type="mRNA"
/db_xref="taxon:31033"
/clone="IMAGE:6353269"
/sex="female"
/tissue type="skin"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="Fugu U7 adult skin"
/notes="Vector: pBluescript-FL; Site 1: PFIM I
(CCANNNTGG); Site 2: PFIM I (CCANNNTGG); BamHI-SmaI
sites were converted to BamHI-PfIMI-SfiI-PfIMI sites (SmaI
is destroyed). Other part of the vector is untouched. The
cDNA is inserted between two PFIM I sites in T3 (5') to T7
(3') direction. Library materials provided by G. Elgar (UK
MRC HMP-RC) and constructed and donated by Drs. K.
Kawakami, M. Sasaki, S. Sugano, K. Kikuchi and S. Watabe
(University of Tokyo, Institute of Medical Science and
Laboratory of Aquatic Molecular Biology and
Biotechnology)."

ORIGIN

Alignment Scores:
Pred. No.: 4.03e-34 Length: 819
Score: 559.50 Matches: 111
Percent Similarity: 56.52% Conservative: 32
Best Local Similarity: 43.87% Mismatches: 99
Query Match: 28.65% Indels: 11
DB: 14 Gaps: 3

US-10-037-417-46 (1-357) x CAS58481 (1-819)

QY 38 CysGlyArgProGluProSerAlaAaGleValGlyGlySerAsnAlaGlnProGlyThr 57
DB 94 TGTGGAAAGACTCAGATCAACAGCAGATGCTGGGTGAGCAGGTGGCCCGAGGCGAGC 153
QY 58 TrpProTrpGlnValSerLeuHisHisGlyGlyHisLeCysGlyGlySerLeuLeu 77
DB 154 TGGCCCTGGCAGGCGCAGTCTGCGAGTCTCTGGGGGTCAACGCTGTGAGGCTGCTCATC 213
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
DB 214 AACAAACGATGGGTGCTGAGCGCAGCCACTGCTTTTCAG-----GGCGTCAGACCC 264
QY 98 AlaAlaGluTrpSerValLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
DB 265 AGCAGCGTG---ACGGTGTCTGGTGTGTCAGACGAGCGAGGCTCAACCCCAACGAG 321
QY 118 HisThrArgAlaValAlaAlaValValProAlaHisValSerGlnValGluLeuGly 137
DB 322 GAGCGGTGAGAGTACACAGATCATCAATCCGACTCACTACAAATACGATTAAC 381
QY 138 AlaAspLeuAlaLeuGluGluLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
DB 382 AACGACATCTCCCTGCTGCACTGGCAGAGCGGTGCTTTTACAACTACATCCAGCGC 441
QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
DB 442 GTCTGCTTGGCGGCGGCGAGGACCTTCCACGGGCGACCGACTCTGGGTACCGCGC 501
QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
DB 502 TGGGGCAACATGGATCGGAGTGGCTCTTCCCTTCCCGGCGGCGGCGGAGTGGAG 561
QY 198 LeuArgLeuGlyGluAlaThrCysGlnCysLeuTySerGlnProGlyProPheAsn 217
DB 561

DB 562 GTGCCCATCAGGGGGAACAGGGAGTGTAACTGTAACTACGGA----- 603
QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
DB 604 ---GTGGGAAGGATTAACGACACATGCTGTGTCGGGTGGCGGAGCGGAGCAAGAT 660
QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGln 257
DB 661 TCTTGTCTCAGGGGGAATTCGGAGGCCCTTTGGTGATCAAGACAGACAGTCGCTGATCCAG 720
QY 258 AlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
DB 721 GCGGGGATCGTGAGTTTCGGACCGCGCTGTGTCAACCTGATACCTACTCCAGCGGTGTACGCT 780
QY 278 AlaValAlaThrTyrGluAlaTrpIleArgGluGlnVal 290
DB 781 CCGGTCTCCAGTATAAACCCTGGATCAACAGTCATC 819

RESULT 10

BC034294

LOCUS

DEFINITION

BC034294

ACCESSION

BC034294.1

VERSION

GI:21706706

KEYWORDS

HTC.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1188)

REFERENCE

1 (bases 1 to 1188)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Udell, T.B., Toshyuki, S.,

Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,

Abrahamson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,

Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, M.A.,

Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

JOURNAL

2 (bases 1 to 1188)

PUBLISHED

Strausberg, R.

REFERENCE

Direct Submission

Submitted (02-JUL-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs@mail.nih.gov

Tissue Procurement: James Cleaver, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,

Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,

A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 52 Row: d Column: 14
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13994275
 This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
 1..1188
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4752547"
 /tissue_type="Skin, squamous cell carcinoma"
 /clone_lib="NCI CGAP_Skn4"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6.1"

ORIGIN

Alignment Scores:
 Pred. No.: 1.5e-33 Length: 1188
 Score: 555.50 Matches: 128
 Percent Similarity: 50.75% Conservative: 32
 Best Local Similarity: 40.63% Mismatches: 118
 Query Match: 28.44% Indels: 37
 DB: 11 Gaps: 10

US-10-037-417-46 (1-357) x BC034294 (1-1188)

QY 32 AlaArgGlyProProTyrCysGlyArgProGluProSerAlaAlaGlyValGlyGlySer 51
 |||
 |||
 Db 113 GCCAAGGAGCAGCAACAGCCTGTGTGCTGCCAGGATGCTGAACCAATGTTGGCGGCGAG 172
 QY 52 AsnAlaGlnProGlyThrTrpProTyrGlnValSerLeuHisHisGlyGlyHisIle 71
 |||
 |||
 Db 173 GACACGAGGAGGCGAGTGTGCTGCAAGTCAAGTCCAGCCGCAACCGGAAGCCATTC 232
 QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91
 |||
 |||
 Db 233 TGGGGGGGACCTCATGCGGAGCAGTGGTCTGAGCGGTGGCGACTGTTCTC----- 286
 QY 92 AsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAsp 111
 |||
 |||
 Db 287 CGCAACACCTCTGAGACGCTCCCTGCTGAGCGTCTGCTGGCGGCAAGCAGCTAGTG 343
 QY 112 GlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValProAlaAsn--- 130
 |||
 |||
 Db 344 CAGCCG-----GGACCAACACGCTATGATGCCGGGGTGGAGCGGTGGAGCAACCC 397
 QY 131 ---TyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuA-gLeuAlaSerProAla 149
 |||
 |||
 Db 398 CTGTACCAAGGCACGGCTCCAGCGCTGACCTGGCCCTGGTGGAGCTGGAGCAACCACTG 457
 QY 150 SerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheValHis 169
 |||
 |||
 Db 458 CCCTTCCCAATTACATCCTCCCGTGTGCTGCTGACCCCTCGGTGATCTTTGAGACG 517
 QY 170 GlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeu 189
 |||
 |||
 Db 518 GGCATGAACCTCTGGGTCACTGGCTGGGCGAGCCCCAGTGGAGGAAGCACTCTGCCCGAA 577
 QY 190 ProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeu 209
 |||
 |||
 Db 578 CCGGGATCCTGCAGAACTCGCTGTGCCCATCATCGACACACCAAGTGCACCTGTCTC 637
 QY 210 TyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGly---MetLeuCys 228
 |||
 |||
 Db 638 TACAGCAAGACACCGAGTTTGGCTTACCAACCCCAAAACCATCAAGATGATGCTGTGC 697
 QY 229 AlaGlyTyrProGluGlyArgAspThrCysGlnGlyAspSerGlyGlyProLeuVal 248
 |||
 |||
 Db 698 GCCGGCTTCGAGGGGCAAGAGATGCTGCAAGGGGCGAGCTCGGGGGGCCCCCTGTGTG 757

QY 249 CysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGly 268
 |||
 |||
 Db 758 TGCTCGTGGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 817
 QY 269 ArgArgAsnArgProGlyValPhe----- 276
 |||
 |||
 Db 818 CGCCAGAACCGCCAGGTGTCTACATCCGTGTCACTCCCGCCACCACTGGATCCATCG 877
 QY 277 -----ThrAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGly 292
 |||
 |||
 Db 878 GATCATCCCAAACTGCAGTCCAGCCAGCGAGGTTGGG---CGCCAGAGTG----- 928
 QY 293 SerGluProGlyProAlaPhePro-----ThrGlnProGlnLysThrGln 307
 |||
 |||
 Db 929 AGACCCCGGGGCGAGGAGCCCTTGAGCAGAGCTGACCCAGCTGCGCGGCCCA--- 985
 QY 308 SerAspCysLeuHisGlnThrAlaPheLeuAspSerAlaArgIle 322
 |||
 |||
 Db 986 -----CACCATCTGCTGCTGCTCTCCAGCGCTGCTGT 1018
 RESULT 11
 CF238383
 LOCUS
 DEFINITION
 AGENCOURT 15097797 NICHED XGC Emb6 Silurana tropicalis cDNA clone
 IMAGE:6993385 5', mRNA sequence.
 ACCESSION
 CF238383
 VERSION
 CF238383.1 GI:33441591
 KEYWORDS
 EST.
 SOURCE
 Silurana tropicalis (western clawed frog)
 ORGANISM
 Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopodinae; Silurana.
 REFERENCE
 1 (bases 1 to 938)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Robert M. Grainger
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAMA14673 row: 1 column: 24
 High quality sequence stop: 724.
 FEATURES
 source
 1..938
 Location/Qualifiers
 /organism="Silurana tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="IMAGE:6993385"
 /tissue_type="neural"
 /dev_stage="embryo, stages 14-19"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NICHED XGC Emb6"
 /note="Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV;
 Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 2.1 kb. Constructed by Invitrogen. Note: This is a
 Xenopus Gene Collection (XGC) library."
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.35e-33 Length: 938
 Score: 551.00 Matches: 113
 Percent Similarity: 55.09% Conservative: 44
 Best Local Similarity: 39.65% Mismatches: 103

```
Query Match: 28.21% Indels: 25
DB: 14 Gaps: 7
US-10-037-417-46 (1-357) x CF238383 (1-938)
QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
118 TCGGGGGTCCCAATAGTAGTCCAAATCGGATAGTAGGGGAACGAACAGTATGAAGGGGAA 177
QY 58 TrpProTrpGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIle 77
178 TGGCCCTGGCAGTCAGCTGAGCTACAAAGGCTCAAACTGTCTGCGGAGGAGCCTTATC 237
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
238 ACCGACAGCTGGGTCTGACTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 237
QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
286 GTTCTCTCAATACATTTTACCTTGAGTTTAT--CAACTGTCTAACCTTAAAAACCCG 342
QY 118 HisThr-----ArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGlu 135
343 AATACGGTATCTAGTGGGTGAAAGAAATCATCATATAAACAGGCTTACCAATACAGGGA 402
QY 136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
403 TCTAGGGGGACATTTGCTTGTGATGACTGGAGAACCGTCACTTCACTTCACTTCACTC 462
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
463 CTCCTAGTCTGCTCCCGCCCGGCTTCTGAACTTCTTCTGCGAGGACCAATATCTGGGT 522
QY 176 ThrGlyTrpGlyAspValGlnGluAlaaspProLeuProLeuProTrpValLeuGlnGlu 195
523 ACTGGATGGGAGATATCAAGAGAGCCAGGATCTTAAGTAACCGGAAACCTTACAGAG 582
QY 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSer----- 211
583 GCTTCGGTGAAGCTAATTGACTGGAATAGCTGTGAGTCCATGTATGAAACCACTTCGGA 642
QY 212 ---GlnProGly---ProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAla 229
643 TACAACACCTAGTGTCTCTTTT-----ATCATGACACACATGTCTGTGTCT 687
QY 230 GlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCys 249
688 GGATACAGAGGAGACAGATAGATGCTGTACGGAGACTCTGTGGGGCTCTTGTATGC 747
QY 250 GluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArg 269
748 AATGTCAACACACGCTGGTGGCAGTATGGCATCATCAGCTGGGGTATTGGATGTGCTGAA 807
QY 270 ArgAsnArgPro-GlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGluGl 289
808 GCCAACGCAACCGGGGGTCTATACCAAGTCCAGTATTATGACTCTTGTGATCAACAGTA 867
QY 289 nValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGlnSerAs 309
868 CAT-----CCTTTCATTAATAATCAGCAAGCAGGATCAGACTGG 909
QY 309 pCysLeuHisGln 313
910 TTGTGATACGAA 922
RESULT 12
BX860661 777 bp mRNA linear EST 16-DEC-2003
LOCUS
DEFINITION BX860661 Rainbow trout multi-tissues (tcba) Oncorhynchus mykiss
cDNA clone tcba0009c.j.09 5prim, mRNA sequence.
ACCESSION BX860661
VERSION BX860661.1 GI:39958625
KEYWORDS EST.
```

```
SOURCE Oncorhynchus mykiss (rainbow trout)
ORGANISM Oncorhynchus mykiss
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
1 (bases 1 to 777)
GOVERNOR M., Guiguen, Y. and Le Gac, F.
Construction and primary characterization of normalized cDNA
libraries in rainbow trout, Oncorhynchus mykiss
Unpublished (2003)
CONTACT: Guiguen Y
INRA - SCRIBE
Campus de Beaulieu, RENNES cedex, 35042, France
Tel: 02.23.48.50.09
Fax: 02.23.48.50.20
Email: Yann.Guiguen@beaulieu.rennes.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at signasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.
Plate: 0009 row: j column: 9
Seq primer: M13R.
FEATURES
Location/Qualifiers
1..777
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcba0009c.j.09"
/tissue_type="multi-tissues"
/dev_stage="from embryos to adults"
/lab_host="DH10B"
/clone_lib="Rainbow trout multi-tissues (tcba)"
/note="Vector: p7T3D-pac; Clone distribution: AGENAE
Resource centre: Francois PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LREG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"
ORIGIN
Alignment Scores:
Pred. No.: 2,11e-33 Length: 777
Score: 550.00 Matches: 108
Percent Similarity: 56.30% Conservative: 35
Best Local Similarity: 42.52% Mismatches: 95
Query Match: 28.16% Indels: 16
DB: 13 Gaps: 4
US-10-037-417-46 (1-357) x BX860661 (1-777)
QY 46 ArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHis 65
23 AAGATTGTGGGGGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 82
QY 66 HisGlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAla 85
83 CGCTCAACCGCCATTTTGTGGAGGCTCTCTATCAACAAGAGTGGTGTGCTGATGCC 142
QY 86 AlaHisCysPhe-----MetThrAsnGlyThrLeuGluProAlaAlaGluTrpSer 102
143 GCTCACTGCTTCCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 181
QY 103 ValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaVal 122
182 GTCTACTCTGGCGGCGGAGAGATCAACATGAGTATGCCAAGCGGTG 241
QY 123 AlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeu 142
242 TCTCAGATCTCTGCAACCCGGAACCTACAGTCAACGACGACGACGACGACGACGACG 301
143 LeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeuProArg 162
302 CTGAAGCTCTCATCCTTCTCCTTCTCCTTCTCCTTCTCCTTCTCCTTCTCCTTCTC 361
```


Db 1660 ACAGAGGGCAACCTCAGTGTGACGGAGAGGAGCTGTAGCAGCGCTCAGATGAGAAG 1719

QY 44 -----SerAlaArgIleValGlySerAsn 52

Db 1720 GACTGGAGCTGTGGCTGGCTGATTCACGACAGAGCTGTGTGTGGGGGCGAGAT 1779

QY 53 AlaGlnProGlyThrProTrpGlnValSerLeuHisHis---GlyGlyGlyHisIle 71

Db 1780 GCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCTGTGCTGTGGGCCGAGGCCACATC 1839

QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91

Db 1840 TCGGGTGTCTCCCTCATCTCTCCCACTGCGTGTCTCTGCGCACACTGCTACATCGAT 1899

QY 92 AsnGly-----ThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSer 109

Db 1900 GACAGAGGATTCAGGTACTCAGACCCCGAGTGGAGCGCTTCTGTGGCTTGCACGAC 1959

QY 110 Gln---AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValValPro 128

Db 1960 CAGAGCCAGCGAGCGCCCTGGGTGAGGAGCGCGAGCTCAAGCGCATCATCTCCAC 2019

QY 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerPro 148

Db 2020 CCTCTTTCATGACTTCACTTCGACTATGACATCGCGCTCTCTGGAGCTGGAGAACCG 2079

QY 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168

Db 2080 GAGAGTACAGCTCCATGTGTGGGCCCATCTGCTGCGGAGCGCTCCATGTCTTCCCT 2139

QY 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuPro 188

Db 2140 GCGCGCAAGGCCATCTGGGTCAAGGCTGGGACACACCCAGTATGAGGCGCTGGCGG 2199

QY 189 LeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208

Db 2200 CTG-----ATCCTGCAAAAGGTCAGATCCCGCTCATCAACACGACACCTTCGCGAAG 2253

QY 209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCys 228

Db 2254 CTCTGCGCGAG-----CAGATCACCGCGCGATGATGTC 2289

QY 229 AlaGlyTrpProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeu--- 247

Db 2230 GTGGCTTCTCCTCAGCGGCGCGCTGAGACTCCTCCAGGGTGTATCCGGGGGACCCCTGTCC 2349

QY 248 ValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCys 267

Db 2350 AGCTGGAGGCGGATGGGCGGATCTTCAGGCGGTGTGTGAGTGGGAGACGGCTGC 2409

QY 268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArg 287

Db 2410 GCTCAGAGGAACAGCGCGCTGTACACAGGCTCCCTCTCTTTCGGGACTGGATCAA 2469

QY 288 Glu 288

Db 2470 GAG 2472

RESULT 15

LOCUS B0729181

DEFINITION AGENCOURT_8217490_NICHD_XGC_Emb2_Xenopus laevis cDNA clone

ACCESSION B0729181

VERSION B0729181.1

KEYWORDS EST.

SOURCE Xenopus laevis (African clawed frog)

ORGANISM Xenopus laevis

REFERENCE 1. (bases 1 to 905)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Igor Dawid
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Cloned through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10808 row: 1 column: 22
High quality sequence stop: 691.

FEATURES
source
1..905
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:4889637"
/tissue_type="embryo, stage 17/19"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD XGC Emb2"
/note="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.1 kb. Constructed by Life Technologies."

ORIGIN

Alignment Scores:
Pred. No.: 5,95e-32 Length: 905
Score: 533.00 Matches: 116
Percent Similarity: 55.90% Conservative: 45
Best Local Similarity: 40.28% Mismatches: 99
Query Match: 27.29% Indels: 29
DB: 13 Gaps: 7

US-10-037-417-46 (1-357) x BQ729181 (1-905)

QY 10 GlyGlnLeuGlyAlaValAlaAsnSerAspSerTyrSerLeuTyrGlyLeuValProSer 29

Db 90 GGGTCTGAGGAGCAGCATCGCTATCAAGTTCTTGAGAAACATATGCTCTA----- 140

QY 30 GlyProAlaArgGlyProProTyrCysGlyArgProGluProSerAlaArgIleValGly 49

Db 141 -----TCAGTCTGCGGTGCGCCCAATGGTATCAGACGAGATAGTAGG 182

QY 50 GlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisGlyGly 69

Db 183 GGAACGGACAGTAAGAGGGGAATGGCCCTGGCAGATCAGTCTGAGCTACAAAGGGAA 242

QY 70 HisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPhe 89

Db 243 CTGTCTGCGGGGATCCCTTATTGCCAACAGTGGATTCGACTGCTGCTCAGTCTT 302

QY 90 MetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuGlyValHisSer 109

Db 303 GATTCG-----CAAAACGTTTCCCAATATAAGTTTACCTGGGGTTTAT--- 347

QY 110 GlnAspGlyProLeuAspGlyAlaHisThr-----ArgAlaValAlaAlaIleVal 127

Db 348 CGATGTCTCTCTCCATAAATCCCAACACAGCTCCAGGAGTGTGAAAGAAATCATATA 407

QY 128 ProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSer 147

Db 408 CACCAGATTACCAATTGAAGGATCTAATGGGACATTCGACCTGATTGAATGGATCAA 467

QY 148 ProAlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPhe 167

Db 468 CTGTGCTACCTTTACCCCTTACATCTCCACCGCTCCGCCACCCAGCTGCTGCTT 527

QY 168 ValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeu 187

Db 528 CTGCGGGGTGAATGCTGGTGAATGCTGGTGAATGCTGGTGAATGCTGGTGAATGCTGCT 587

```

QY 188 ProLeuProTyrValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGln 207
Db 588 AGTAACCCCTAAACCCCTGCAGAGGCTACAGTGAAGTCTAATGACTGGCATAGCTGTGAG 647
QY 208 CysLeuTyrSer-----GlnProGly---ProPheAsnLeuThrLeuGln 221
Db 648 TCCATGTACGAAACCCAGCTTTGGGATACAAACCTAATGTCCTTTT----- 692
QY 222 IleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgAspThrCysGlnGly 241
Db 693 ATCCTGGATCACATGTTCTGTGCTGGCTCAAGGAGGAAATAGACGCCCTGTCTAGGGA 752
QY 242 AspSerGlyGlyProLeuValCysGluGluGlyGlyArgTyrPheGlnAlaGlyIleThr 261
Db 753 GACTCTGGAGGACCTCTTGTATGCAGAGTCAACAACACATGTTGGCAATACGGCATTTGTC 812
QY 262 SerPhe-GlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAlaTh 281
Db 813 AGCTGGGGGAATTGGATGTGGACAGCTAACCAACCGGT-GGCTATATACCAAGGTCCAGTA 871
QY 281 rTyrGluAlaIleArgGlu 288
Db 872 TTTGGAGCTTGGATCAAGCAG 893

```

Search completed: February 25, 2004, 18:01:10
Job time : 2501 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 15:58:04 ; Search time 96 Seconds
(without alignments)

2063.724 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQKGLGPGQLGAVANSBS

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp
-O=/cgn2_1/uspto/spool_p/us10037417/runat_24022004_101039_17477/app_query.fasta_1.519
-DB=Issued_Patents_NA -QFMT=fastp -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10037417 @CGN 1 1 103 @runat_24022004_101039_17477 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPHLOCK=100 -LONGLOG
-DRV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	603.5	30.9	1613	4	US-09-387-375-1
2	564.5	28.9	1142	4	US-09-386-642-8
3	563.5	28.9	1110	4	US-09-386-653A-1
4	562.5	28.8	1212	4	US-09-620-312D-431
5	561.5	28.8	980	4	US-09-023-942A-30
6	557	28.5	1130	4	US-09-387-375-8
7	550	28.2	1169	4	US-09-386-642-7
8	549	28.1	3147	2	US-09-027-337-1
9	549	28.1	3147	4	US-09-644-600-1
10	549	28.1	3147	4	US-09-644-600-18
11	549	28.1	3147	4	US-09-654-600A-1
12	549	28.1	3147	4	US-09-654-600A-18

13	545	27.9	1378	4	US-09-907-794A-262	Sequence 262, Appl
14	545	27.9	1378	4	US-09-905-125A-262	Sequence 262, Appl
15	545	27.9	1378	4	US-09-902-775A-262	Sequence 262, Appl
16	544	27.9	1430	4	US-09-386-629-1	Sequence 1, Appli
17	538.5	27.6	1081	3	US-09-008-271A-15	Sequence 15, Appl
18	538.5	27.6	1100	4	US-09-907-794A-256	Sequence 256, Appl
19	538.5	27.6	1100	4	US-09-905-125A-256	Sequence 256, Appl
20	538.5	27.6	1100	4	US-09-902-775A-256	Sequence 256, Appl
21	537.5	27.5	1100	4	US-09-023-942A-5	Sequence 5, Appli
22	532	27.2	2152	4	US-09-023-655-157	Sequence 157, App
23	531	27.2	1130	4	US-09-386-653A-8	Sequence 8, Appli
24	530.5	27.2	1094	4	US-09-023-942A-3	Sequence 3, Appli
25	529	27.1	959	4	US-09-023-942A-25	Sequence 25, Appl
26	514	26.3	933	4	US-09-023-942A-29	Sequence 29, Appl
27	513	26.3	1108	2	US-09-016-366A-14	Sequence 14, Appl
28	509	26.1	1165	4	US-09-023-942A-28	Sequence 28, Appl
29	507	26.0	1153	4	US-09-280-116-10	Sequence 10, Appl
30	507	25.7	1166	4	US-09-386-629-2	Sequence 2, Appli
31	501	25.7	1103	4	US-09-386-642-59	Sequence 59, Appl
32	497.5	25.5	1605	2	US-09-000-846-1	Sequence 1, Appli
33	497.5	25.4	1097	2	US-08-978-404B-4	Sequence 4, Appli
34	496.5	25.4	1103	2	US-09-016-366A-24	Sequence 24, Appl
35	495	25.3	2900	4	US-09-644-600-9	Sequence 9, Appli
36	495	25.3	2900	4	US-09-654-600A-9	Sequence 9, Appli
37	494.5	25.3	1128	2	US-09-016-366A-20	Sequence 20, Appl
38	494.5	25.3	1128	2	US-08-978-404B-15	Sequence 15, Appl
39	494.5	25.3	1137	2	US-09-016-366A-18	Sequence 18, Appl
40	494.5	25.3	1137	2	US-08-978-404B-13	Sequence 13, Appl
41	493.5	25.3	1081	2	US-09-016-366A-22	Sequence 22, Appl
42	493.5	25.3	1081	2	US-08-978-404B-17	Sequence 17, Appl
43	493.5	25.3	1615	4	US-09-820-002-1	Sequence 1, Appli
44	492.5	25.2				
45	492.5	25.2				

ALIGNMENTS

RESULT 1
US-09-387-375-1
; Sequence 1, Application US/09387375
; Patent No. 6485957
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Oj, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/09/387,375
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatenIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-387-375-1

Alignment Scores:	1.7e-40	Length:	1613
Pred. No.:	603.50	Matches:	139
Score:		Conservative:	38
Percent Similarity:	50.00%	Mismatches:	129
Best Local Similarity:	39.27%	Indels:	48
Query Match:	30.90%	Gaps:	7
DB:	4		
US-10-037-417-46 (1-357) x US-09-387-375-1 (1-1613)			
QY	38	CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr	57
Db	150	TCCGGGGACCCCGCATTCGATCGTTGGGGCCCGCGGATGGCCGACGAGAG	209
QY	58	TrpProTrpGlnValSerLeuHisGlyGlyHisIleCysGlyGlySerLeuIle	77

Db 653 TTGTCACAGAGACATGTTGTGTGCTATGTGAGGGGGCAAGGACGCGCTGCCAGG 712
 QY 241 lYAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleT 261
 Db 713 GTGACTCTGGGGCCCACTCTCTGCTGCTGTGAGGGTCTCTGTATACCTGACGGGCATTG 772
 QY 261 hrSerPheGlyPheGlyCysGlyArgGAsnArgProGlyValPheThrAlaValAlaT 281
 Db 773 TGAGCTGGGGAGATGCTGTGGGGCCGCAACAGGCTGTGTGTATACACTCTGTGGCTCCA 832
 QY 281 hrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyProAlaPheProT 301
 Db 833 GCTATGCTCTGATGATCAAGCAAGGTG-----ACAGAACTCCAGCCTCTGTGTGTGTC 886
 QY 301 hrGlnProGlnLysThrGlnSerAsp-----CysLeuHisGlnThrAlaPhe 316
 Db 887 CCNAAACCCAGGATGCCACCCGACGAACTCTGTGGCAGCCACCTGGGCTTC 942
 RESULT 3
 US-09-386-653A-1
 ; Sequence 1, Application US/09386653A
 ; Patent No. 6458564
 ; GENERAL INFORMATION:
 ; APPLICANT: Andrade-Gordon, Patricia
 ; APPLICANT: Darrow, Andrew
 ; APPLICANT: Qi, Jian-shen
 ; TITLE OF INVENTION: DNA encoding the novel human serine
 ; TITLE OF INVENTION: protease T
 ; FILE REFERENCES: ORT-1032
 ; CURRENT APPLICATION NUMBER: US/09/386,653A.
 ; CURRENT FILING DATE: 1999-08-31
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1110
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-386-653A-1

Alignment Scores: 1.97e-37 Length: 1110
 Pred. No.: 563.50 Matches: 128
 Score: 51.28% Conservative: 32
 Percent Similarity: 41.03% Mismatches: 121
 Best Local Similarity: 28.85% Indels: 32
 Query Match: 4 Gaps: 9
 Db: 4

US-10-037-417-46 (1-357) x US-09-386-653A-1 (1-1110)

QY 32 AlaArgGlyProProTyrCysGlyArgProGluProSerAlaAlaGlyValGlySer 51
 Db 93 GCCAAGGCAGCAACAGCCTGTGTGTCGCCCCAGGATGCTGAACCGAATGTTGGCGGGCAG 152
 QY 52 AsnAlaGlnProGlyThrTrpProTyrGlnValSerLeuHisGlyGlyHisIle 71
 Db 153 GACACGCGAGGGCGAGTGGCCCTGGCAAGTCCAGCATCCACGCGCAAGCAAGCACTTC 212
 QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91
 Db 213 TGCGGGGGCGACCTCATCGGAGCAGTGGGTCTGACGGCTGCGACATGCTTC----- 266
 QY 92 AsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAsp 111
 Db 267 CGCAACACTCTGAGACCTCCCTG---TACCAGGTCTCTGGGGGCAAGCAGCTAGTG 323
 QY 112 GlyProLeuAspGlyAlaHisThrArgAlaValAlaIleValValProAlaAsn--- 130
 Db 324 CAGCGG-----GGACCACACGCTATGTCGCGGGTGAGCGAGCTGGAGCAACACCC 377
 QY 131 ---TyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAla 149
 Db 378 CTGTACCAAGGCAGCGGCTCCAGCGCTGACGTGGCCCTGGTGGAGGCGACCACTGT 437

QY 150 SerLeuGlyProAlaValTrpProValCysLeuProAlaSerHisArgPheValHis 169
 Db 438 CCCTTCACCAATTACATCTCCCGGTGTGCTGCTGACCCCTCGGTGATCTTTGAGACG 497
 QY 170 GlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeu 189
 Db 498 GGCATGAACCTGTGGGTCACTGCTGGGGCAGCCCGCAGTGGAGAAACCTCTCTGCCCGAA 557
 QY 190 ProTrpValLeuGlnValLeuArgLeuGlyGluAlaThrCysGlnCysLeu 209
 Db 558 CCGCGGATCTCGAGAAACTCCTGTGCCCATCATCGACACACCCCAAGTGCACCTGCTC 617
 QY 210 TyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGly---MetLeuCys 228
 Db 618 TACAGCAAAAGACACCGAGTTTGGCTACCAACCAACCAACCATCAAGAATGACATGCTGTGC 677
 QY 229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuVal 248
 Db 678 GCGCGCTTCGAGGAGGCGAAGAGATGCTGCAAGGCGACTCGGGCGGCGCCCTGGTG 737
 QY 249 CysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGly 268
 Db 738 TGGCTCGTGGTCACTGCTGGCTGAGGGGGGTGATCAGCTGGGTGAGGGCTGTGCC 797
 QY 269 ArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGlu 288
 Db 798 CGCCAGAACCGCCAGGTGTCTATCATCCGTGTACCGCCGCCACCACTGGATCCATCGG 857
 QY 289 -----GlnValMetGlySerGlu-----Pro 295
 Db 858 ATCATCCCAAACTGCAGTTCCA-GCCAGCGAGTGTGGCGGCGCAGAGTGAGACCCCGG 916
 QY 296 GlyProAlaPhePro-----ThrGlnProGlnLysThrGlnSerAspCys 310
 Db 917 GGGCCAGGAGGCCCTTGAGCAGAGCTCTGCACCCAGCGCTGCCCGCCCA----- 964
 QY 311 LeuHisGlnThrAlaPheLeuAspSerAlaArgIle 322
 Db 965 ---CACCATCTGCTGTGCTCTCCACGCGCTGTGTT 997
 RESULT 4
 US-09-620-312D-431
 ; Sequence 431, Application US/09620312D
 ; Patent No. 6569662
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Liu, Chonghua
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Chen, Rui-hong
 ; APPLICANT: Zhao, Qing A.
 ; APPLICANT: Wehrman, Tom
 ; APPLICANT: Xue, Aidong J.
 ; APPLICANT: Yang, Yonghong
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Zhou, Ping
 ; APPLICANT: Ma, Yungqing
 ; APPLICANT: Wang, Dunrui
 ; APPLICANT: Wang, Zhiwei
 ; APPLICANT: John Tillinghast
 ; APPLICANT: Drmanac, Radoje T.
 ; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
 ; TITLE OF INVENTION: Polypeptides
 ; FILE REFERENCE: 784CIP2B
 ; CURRENT APPLICATION NUMBER: US/09/620,312D
 ; CURRENT FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 09/552,317
 ; PRIOR FILING DATE: 2000-04-25
 ; PRIOR APPLICATION NUMBER: 09/488,725
 ; PRIOR FILING DATE: 2000-01-21
 ; NUMBER OF SEQ ID NOS: 1105
 ; SOFTWARE: pt_FL_genes version 1.0

```

; SEQ ID NO 431
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (135)..(1007)
US-09-620-312D-431

Alignment Scores:
Pred. No.:      2,69e-37      Length:      1212
Score:          562.50       Matches:    128
Percent Similarity: 51.26%   Conservative: 32
Best Local Similarity: 41.03% Mismatches:    121
Query Match:     28.80%     Indels:      32
DB:              Gaps:        9

US-10-037-417-46 (1-357) x US-09-620-312D-431 (1-1212)

QY 32 AlaArgGlyProTyrCysGlyArgProGluProSerAlaAArgIleValcylglySer 51
Db 192 GCACAGCGAGCAACAGCTGTGGTGCCGCCAGGATGTCACCGCATGTGGCGGGCAG 251
QY 52 AsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisGlyGlyHisIle 71
Db 252 GACACGACGAGCGGCAGTGGCCCTGGCAAGTCAGCATCCAGCGCACGAAGCACTTC 311
QY 72 CysGlyGlySerLeuIleAlaProSerTrpValSerAlaAHisCysPheMetThr 91
Db 312 TGGGGGGGAGCCTCATCCGAGCAGATGGTGGTCCGAGCGCTGCCCATCTGTTTC----- 365
QY 92 AsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAsp 111
Db 366 CGCAACACTCTGAGACGTCCTG---TACCAGTCTCTGCTGGGGGCAAGCAGCAGTAGTG 422
QY 112 GlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsn--- 130
Db 423 CAGCCG----GGACCAACAGCTATGTATGCCGGGTGAGCGAGTGGAGAGCAACCCC 476
QY 131 ---TyrSerGlnValGluLeuGlyAlaAapLeuAlaLeuArgLeuAlaSerProAla 149
Db 477 CTGTACAGGCGCACGGCTCCAGCGCTGAGTGGCCCTGGTGGAGCTGGAGCACCACTG 536
QY 150 SerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheValHis 169
Db 537 CCCTTACCAATTACATCTCCCGTGTGCTGCTGACCCCTCGGTGATCTTTGAGAGC 596
QY 170 GlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeu 189
Db 597 GGCAATGAACCTCTGGGTCACTGGCTGGGCGAGCCCCAGTGAAGAAGCACTCTGCCCGAA 656
QY 190 ProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeu 209
Db 657 CCGCGGATCTGCAGAACACTCGTGTGCCCATCATCGACACACCCCAAGTGCACTGTGCT 716
QY 210 TyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGly---MetLeuCys 228
Db 717 TACAGCAAGACACCGATTTGGCTACCAACCCCAACCATCAAGAATGACATGCTGTGC 776
QY 229 AlaGlyTrpProGluGlyArgAspThrCysGlnGlyAspSerGlyGlyProLeuVal 248
Db 777 GCCGGCTTCGAGGAGGCGCAAGAGATGCTGCAAGGGCGACTCGGGCGGCCCTCTGGTG 836
QY 249 CysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGly 268
Db 837 TGCTCTGGGTGTCAGTCTGGTCTGCAGCGGGGGGTATCAGCTGGGGTGGGGCTGTGCC 896
QY 269 ArgArgAsnArgProGlyValPheThrAlaValAlaAlaThrTyrGluAlaTrpIleArgGlu 288
Db 897 GCCAGAACCCCGCAGGTGTTTACATCGTGTGTACCGGCTACCGCCCAACCACTGATCCATCG 956
QY 289 -----GlnValMetGlySerGlu-----Pro 295

```



```
QY 58 TrpProTyrGlnValSerLeuHisHisGlyGlyGlyHisGlyHisGlySerLeuLeu 77
Db 63 TGGCCCTGGCAGTACAGTACAGCCACAGGAGGACCACTTCTGGGGGGGAGCCTCATC 122
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
Db 123 GCGGAGCAGTGGTCTCTGAGCGGTGGCGACTGCTTC-----CGCAACACCTCGAGAGC 176
QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
Db 177 TCCTCTG---TACCAGTCTCTGCTGGGGGCAAGCAGCTAGTCAGCGC-----GGACCA 227
QY 118 HisThrArgAlaValAlaAlaAlaValValProAlaAsn-----TyrSerGlnValGlu 135
Db 228 CAGCTATGTATCCCGGGTGGAGGAGTGGAGGACCAACCCCTGTACAGGGGACGGCC 287
QY 136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 288 TCCAGCGCTGACGTGGCCCTGGTGGAGCTGGAGGACCAAGTGCCTTCACCAATTATC 347
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 348 CTCCTCGGTGCTGCTGACCCCTCGGTGATCTTTGAGACGGGCATGAATCTGCTGTC 407
QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuTrpValLeuGlnGlu 195
Db 408 ACTGGCTGGGGGACCCAGTGGAGAGACTCTCTGCCCGCAACCGCGATCTTCGAGAA 467
QY 196 ValLeuLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db 468 CTCCTGCTGCCATCATCGACACACACCAAGTGCACCTGTCTACAGCAAAAGACACGAG 527
QY 216 PheAsnLeuThrLeuGlnIleLeuProGly---MetLeuCysAlaGlyTyrProGluGly 234
Db 528 TTTGGCTACCAACCAACCAACCATCAGATGACATGCTGTGCGCGGCTTCGAGGAGGC 587
QY 235 ArgArgAspThrCysGlnGlyAspSerGlyCysLeuValCysGluGluGlyGlyArg 254
Db 588 ANGAAGATGCTGCAGAGGAGTCTGGGGGGGCCCCCTGGTGTGCTCGTGGGTCACTCG 647
QY 255 TrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGly 274
Db 648 TGGCTGACAGCGGGGTGATCAGCTGGGGTGGAGGCTGTGCGCGGACCAACCGCCAGGT 707
QY 275 ValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGlu-----Gln 289
Db 708 GTCTACATCCGTGTACCGCCACACCAACTGGATCCATCGGATCATCCCAACTCGAG 767
QY 290 ValMetGlySerGlu-----ProGlyProAlaPhePro--- 300
Db 768 TTCCA-GCCACGAGGTGGGGGCGGACAGTGGAGACCCCGGGGCCAGGAGCCCTTGA 826
QY 301 -----ThrGlnProGlnIleThrGlnSerAspCysLeuHisGlnThrAlaPhe 316
Db 827 GCAGAGCTCTGCACCCAGCCTGCCGCCCA-----CACCATCTCTGCTGGT 871
QY 317 LeuAspSerAlaArgIle 322
Db 872 CTTCCACGCGCTGCTGT 889
RESULT 6
; Sequence 8, Application US/09387375
; Patent No. 6485957
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/09/387,375
```

```
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1130
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid
; OTHER INFORMATION: sequence of EOS zymogen fusion gene
US-09-387-375-8

Alignment Scores:
Pred. No.: 6,91e-37 Length: 1130
Score: 557.00 Matches: 114
Percent Similarity: 56.63% Conservative: 27
Best Local Similarity: 45.78% Mismatches: 98
Query Match: 28.52% Indels: 10
DB: 4 Gaps: 3

US-10-037-417-46 (1-357) x US-09-387-375-8 (1-1130)
QY 46 ArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTyrGlnValSerLeuHis 65
Db 163 AGATCGTTGGGGCTATGCTCTAGAGGACGAGAGTGGCGGTGGCAGCGGAGCATCCAG 222
QY 66 HisGlyGlyGlyHisIleCysGlyGlySerLeuLeuAlaProSerTrpValLeuSerAla 85
Db 223 CATCTGGGGGACACACGTGTGGGGGGTGGCTCATGCCCGCCCGGCTGGGTGTGACAGCG 282
QY 86 AlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeu 105
Db 283 GCACATGCTTCCCGAGGAGGGCAGCTG-----CCAGCTGAGTACCGCGTGGCGCTG 333
QY 106 GlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAla 125
Db 334 GGGCGCTGGCTGGGGTCCACCTCGCCCGGACGCTCTCGGTGGCGGTGGCGAGGGTG 393
QY 126 ValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeu 145
Db 394 CTGCTGCCCGGACTACTCCGAGGACGGGGCGCGCGGCGGACCTGGCACTGTCTGCACTG 453
QY 146 AlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHis 165
Db 454 CGTGGCGCGGTGGCGCTGAGCGCTCGGCTCCAAACCGCTGTCTGCTGCGCGTGGCGGCGCC 513
QY 166 ArgPheValHisGlyThrAlaCysTrpAlaThrGlyAspValGlnGluAlaAsp 185
Db 514 CGCCCGCGCGCGGACACCATCGCGGTCCCGGCTGGGGGAGCTCCGCCAGGAGTG 573
QY 186 ProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeuGlyGluAlaThr 205
Db 574 CCGCTCCAGAGTGGCGACCGCTACAGAGAGTAAAGGTGCGGTGCTGGTGGAGCTCGCGGACC 633
QY 206 CysGlnCysLeuTyr-----SerGlnProGlyProPheAsnLeuThrLeuGln 221
Db 634 TGGAGCGGCTCTACACAGCTGGCGCGGAGCTGCCCGGAGCTGAGCGCATI----- 684
QY 222 IleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGlnGly 241
Db 685 GTGCTGCTGGAGTCTGTGTGCGCGCTACCCCGGCGGCGGCGGAGGAGCGCTGCGGAGGT 744
QY 242 AspSerGlyProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThr 261
Db 745 GATTCTGGGGACCTCTGACCTGCTGAGTCTGGAGCTGGGTCTGGTGGGGGTGGTGTG 804
QY 262 SerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAlaThr 281
Db 805 AGCTGGGGGAGGGGTGTGGCCCTGCCCAACGCTCCAGGGGTCTACACAGGTGTGGCCACA 864
QY 282 TyrGluAlaTrpIleArgGluGlnVal 290
Db 865 TATAGCCCTGGATTACAGCTCGGCTC 891
```


Thu Feb 26 13:50:26 2004

129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerPro 148
1223 CCCTTCTTCAATGACTTACCTTCGACTATGACATCGCTGCTGGAGCTGGAGAAACCG 2182
149 AlaSerLeuGlyProAlaValTyrProValCysLeuProArgAlaSerHisArgPheVal 168
2183 GCAGATACAGCTCATGCTGGGCGCCATCTGCTGCGGAGCTCCCATGCTTCCCT 2242
169 HisGlyThrAlaCysTyrAlaThrGlyTyrGlyAspValGlnGluAlaAspProLeuPro 188
2243 GCCGCAAGGCCCATCTGGTCAACGGCTGGGAGACACACCCAGTATGGAGGCACTGGCGCG 2302
189 LeuProTyrValLeuGlnValGluLeuArgLeuGlyGluAlaThrCysGlnCys 208
2303 CTG-----ATCCTGCAAAAGGCTGAGATCCGCTGATCAACCAACACCTGGCAGAAC 2356
209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnLeuLeuProGlyMetLeuCys 228
2357 CTCCTGCCGCGAG-----CAGATCACGCGCGCATGATGTC 2392
229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeu--- 247
2393 GGGGCTTCTCAGCGCGCGCTGACTCTCCAGGGGTGATTCGGGGGAGCCCTGCTCC 2452
248 ValCysGluGluGlyGlyArgTyrPheGlnAlaGlyIleThrSerPheGlyPheGlyCys 267
2453 AGCGTGGAGCGGATGGCGGATCTTCCAGGCGGCTGCTGAGTGGGAGGACGGCTGC 2512
268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTyrIleArg 287
2513 GCTCAGAGAACAGCAGCGGTGTACACAGGCTCCCTCTGTTCCGGGACTGGATCAAA 2572
288 GluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
2573 GAGAACACTGGGGTATAGGGCGCGGGGCCA----- 2602
308 SerAspCysLeuHis 312
2603 CCCAAATGTGTACAC 2617

RESULT 9
US-09-644-600-1
; Sequence 1, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-1

Alignment Scores:
Pred. No.: 1,268-35
Score: 549.00
Percent Similarity: 51.80%
Best Local Similarity: 38.69%
Query Match: 28.11%
DB: 4
Length: 3147
Matches: 118
Conservative: 40
Mismatches: 103
Indels: 44
Gaps: 8

US-10-037-417-46 (1-357) x US-09-644-600-1 (1-3147)
32 AlaArgGlyProProTyrCysGlyArgProGluPro----- 43
1763 AGCAAGGCAACCTTGAGTGTGACGGGAAGGAGACTGTAGCGACGGCTCAGATGAGAAG 1822
44 -----SerAlaArgIleValGlyGlySerAsn 52
1823 GACTGCGACTGTGGGCTCGGCTCATTCACGAGACAGGCTGCTGTGTGTGGGGGACCGAT 1882
53 AlaGlnProGlyThrTyrProTyrGlnValSerLeuHisHis-----GlyGlyGlyHisIle 71
1883 GCGGATGAGGGCGAGTGGCCCTGCGAGGTAGCTGCTGCTGCGGCGAGGCGCACATC 1942
72 CysGlyGlySerLeuIleAlaProSerTyrValLeuSerAlaAlaHisCysPheMetThr 91
1943 TGGGCTGCTTCCCTCATCTCTCCCAACTGGTGTCTCTCGGCACACACTGTACATCAT 2002
92 AsnGly-----ThrLeuGluProAlaAlaGluTyrSerValLeuLeuGlyValHisSer 109
2003 GACAGAGGATTCAGGTACTCAGACCCCGAGTGGAGCGGCTTCTGGGCTTGACGAC 2062
110 Gln---AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValPro 128
2063 CAGACCCAGCGCAGCGCCCTGGGTGCGAGGCGCAGGCTCAAGCGCATCATCTCCAC 2122
129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerPro 148
2123 CCCTTCTTCAATGACTTACCTTCGACTATGATCATCGGCTGCTGAGCTGGAGAAACCG 2182
149 AlaSerLeuGlyProAlaValTyrProValCysLeuProArgAlaSerHisArgPheVal 168
2183 GCAGAGTACAGCTCCATGCTGGCGCCATCTGCTGCGGAGCGCTCCCATGCTTCCCT 2242
169 HisGlyThrAlaCysTyrAlaThrGlyTyrGlyAspValGlnGluAlaAspProLeuPro 188
2243 GCCGCAAGGCCCATCTGGGTACCGGCTGGGAGCACACCCAGTATGGAGGCACTGGCGCG 2302
189 LeuProTyrValLeuGlnGluValGluLeuArgLeuGlyGluAlaThrCysGlnCys 208
2303 CTG-----ATCCTGCAAAAGGCTGAGATCCGCTCATCAACCAACACCTCGGAGAAC 2356
209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCys 228
2357 CTCCTGCCGCGAG-----CAGATCACGCGCGCATGATGTC 2392
229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeu--- 247
2393 GTGGGCTTCTCAGCGCGCGGCTGAGTCTCTGCGAGGCTGATTCGGGGGACCCCTGCTCC 2452
248 ValCysGluGluGlyGlyArgTyrPheGlnAlaGlyIleThrSerPheGlyPheGlyCys 267
2453 AGCGTGGAGCGGATGGCGGATCTTCCAGGCGGCTGTGTGAGCTGGGAGACGGCTGC 2512
268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTyrIleArg 287
2513 GCTCAGAGAACAGCAGCGGTGTACACAGGCTCCCTCTGTTCCGGGACTGGATCAAA 2572
288 GluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
2573 GAGAACACTGGGGTATAGGGCGCGGGGCCA----- 2602
308 SerAspCysLeuHis 312
2603 CCCAAATGTGTACAC 2617

RESULT 10
US-09-644-600-18/c
; Sequence 18, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi

```

; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 18
; LENGTH: 3147
; TYPE: RNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Antisense of TADG-15
US-09-644-600-18

Alignment Scores:
Pred. No.: 1,26e-35 Length: 3147
Score: 549.00 Matches: 118
Percent Similarity: 51.80% Conservative: 40
Best Local Similarity: 38.69% Mismatches: 103
Query Match: 28.11% Indels: 44
Gaps: 8
DB:

US-10-037-417-46 (1-357) x US-09-644-600-18 (1-3147)
QY 32 AlaArgGlyProProTyrCysGlyArgProGluPro----- 43
Db 1385 AGCAAGGGCAACCTCAGTGTGACGGAGAGGAGACTGTAGCGACGGCTCAGATGAGAAG 1326
QY 44 -----SerAlaArgIleValGlyGlySerAsn 52
Db 1325 GACTGCGACTGTGGGTGGTGGTCAATTCACAGACAGGCTCGTGTGTGGGGCGACGGAT 1266
QY 53 AlaGlnProGlyThrTrpProTyrGlnValSerLeuHisHis-----GlyGlyGlyHisIle 71
Db 1265 GCGGATGAGGGCGAGTGGCCCTGGCAGGTAAAGCTGCATGCTCTGGCGCCAGGGGCCACATC 1206
QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91
Db 1205 TGCGGTGCTTCCCTCATCTCCCAACTGGCTGTCTGCGGCACACTGCTACATCGAT 1146
QY 92 AsnGly-----ThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSer 109
Db 1145 GACAGAGATTCAGGTACTCAGACCCACCCAGTGGAGCGCTTCTGGGCTTGCAGCAG 1086
QY 110 Gln-----AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaIleValPro 128
Db 1085 CAGAGCCAGCGCAGCGCCCTGGGGTGCAGGAGCGGAGGCTCAAGCGCATCATCTCCAC 1026
QY 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerPro 148
Db 1025 CCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTGCTGGAGCTGGAGAACCG 966
QY 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168
Db 965 GCAGGTACAGCTCCTGCTGGTGGGCCCATCTGCTGCGGAGCGCTCCCATGCTTCCCT 906
QY 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAlaAppProLeuPro 188
Db 905 GCCGCAAGGCCATCTGGGTACCGGCTGCGGACACACCCAGTATGGAGGCACCTGGCGCG 846
QY 189 LeuProTrpValLeuGlnGluValGluLeuArgLeuGlyGluAlaThrCysGlnCys 208
Db 845 CTG-----ATCCTGCAAAAGGTGAGATTCGCGTCATCAACAGACCCACTTGGAGAAC 792
QY 209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCys 228
Db 791 CTCCTGCCGCGAG-----CAGATCACCGCCCGCATGATGTCG 756
QY 229 AlaGlyTyrProGluGlyArgAspThrCysGlnGlyAspSerGlyGlyProLeu----- 247
```

```

Db 755 GTGGGCTTCTCAGCGCGCGGTGGACTCTCTCCAGGGTAITCCGGGGACCCCTGTCC 696
QY 248 ValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCys 267
Db 695 AGGTGGAGCGGATGGCGGATCTTCCAGGCCGGTGTGGTGGAGCTGGGAGACGGCTGC 636
QY 268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArg 287
Db 635 GCTCAGAGGAACAAGCCAGGCGGTGTACACAAGGCTCCCTCTGTTCGGGACTGGATCAA 576
QY 288 GluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
Db 575 GAGAACACTGGGGTATAGGGGCGGGGCCA----- 546
QY 308 SerAspCysLeuHis 312
Db 545 CCCAATGTGTACAC 531

RESULT 11
US-09-654-600A-1
; Sequence 1, Application US/09654600A
; Patent No. 6649741
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirocoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/654,600A
; CURRENT FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-654-600A-1

Alignment Scores:
Pred. No.: 1,26e-35 Length: 3147
Score: 549.00 Matches: 118
Percent Similarity: 51.80% Conservative: 40
Best Local Similarity: 38.69% Mismatches: 103
Query Match: 28.11% Indels: 44
Gaps: 8
DB:

US-10-037-417-46 (1-357) x US-09-654-600A-1 (1-3147)
QY 32 AlaArgGlyProProTyrCysGlyArgProGluPro----- 43
Db 1763 AGCAAGGGCAACCTCAGTGTGACGGAGAGGAGACTGTAGCGACGGCTCAGATGAGAAG 1822
QY 44 -----SerAlaArgIleValGlyGlySerAsn 52
Db 1823 GACTGCGACTGTGGGTGGTGGTCAATTCACAGACAGGCTGCTGTGTGGGGCGACGGAT 1882
QY 53 AlaGlnProGlyThrTrpProTyrGlnValSerLeuHisHis-----GlyGlyGlyHisIle 71
Db 1883 GCGGATGAGGGCGAGTGGCCCTGGCAGGTAGAGCTGCTGCGGCGAGGCCACATC 1942
QY 72 CysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThr 91
Db 1943 TGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGTCTCTGCGGCACACTGCTACATCGAT 2002
QY 92 AsnGly-----ThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSer 109
Db 2003 GACAGAGGATTCAGGTACTCAGACCCACCGAGTGGACGCTTCTCTGGGCTTGCACGAC 2062
```

QY 110 Gln---AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaValValPro 128
 Db 2063 CAGAGCCAGGCGAGCCCTGGGGTGCAGGAGCGAGGCTCAAGCGCATCATCTCCAC 2122
 QY 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerPro 148
 Db 2123 CCCTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTGCTGAGCTGGAGAACCG 2182
 QY 149 AlaSerLeuGlyProAlaValTrrProValCysLeuProArgAlaSerHisArgPheVal 168
 Db 2183 GCAGAGTACAGCTCCATGCTGGCGCCCATCTGCTCGCGAGCGCTCCCATGCTTCCCT 2242
 QY 169 HisGlyThrAlaCysTrrAlaThrGlyTrrGlyAspValGlnGluAlaAspProLeuPro 188
 Db 2243 GCGGCAAGGCCATCTGGGTCAGCGGCTGGGGACACACCAGTATGGAGGCACTGGCGCG 2302
 QY 189 LeuProTrrValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208
 Db 2303 CTG-----ATCCTGCAAAAGGTGAGATCGGCTCATCAACAGACCACCTCGGAGAAC 2356
 QY 209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnLeuProGlyMetLeuCys 228
 Db 2357 CTCTGCGCGAG-----CAGATCAGCGCGCATGATGTC 2392
 QY 229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeu--- 247
 Db 2393 GTGGGCTTCTCAGCGCGCGTGGACTCTCGCGAGGGTGAATCCGGGGAGCCCTCTCC 2452
 QY 248 ValCysGluGluGlyGlyArgTrrPheGlnAlaGlyTrrSerPheGlyPheGlyCys 267
 Db 2453 AGCTGTGAGCGGATGGCGGATCTTCCAGCGCGTGTGGTGTGAGCTGGGAGAGCGGTGC 2512
 QY 268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrrPileArg 287
 Db 2513 GCTCAGAGGAACAAGCAGCGCGTGTACACAGGCTCCCTGTGTTTGGGGACTGGATCAA 2572
 QY 288 GluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
 Db 2573 GAGAACACTGGGGTATAGGGCGCGGGCCA----- 2602
 QY 308 SerAspCysLeuHis 312
 Db 2603 CCCAAATGTGTACAC 2617

RESULT 12

; Sequence 18, Application US/09654600A
 ; Patent No. 6649741
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; APPLICANT: Tanimoto, Hirotochi
 ; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
 ; TITLE OF INVENTION: Overexpressed in Carcinomas
 ; FILE REFERENCE: D6064CIP/D
 ; CURRENT APPLICATION NUMBER: US/09/654,600A
 ; CURRENT FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 09/421,213
 ; 09/027,337
 ; PRIOR FILING DATE: 1999-10-20
 ; 1998-02-20
 ; NUMBER OF SEQ ID NOS: 98
 ; SEQ ID NO 18
 ; LENGTH: 3147
 ; TYPE: RNA
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: Antisense of TADG-15
 US-09-654-600A-18

Alignment Scores: 1.26e-35 Length: 3147
 Pred. No.: 549.00 Matches: 118
 Score:

Percent Similarity: 51.80% Conservative: 40
 Best Local Similarity: 38.69% Mismatches: 103
 Query Match: 28.11% Indels: 44
 Db: 4 Gaps: 8
 US-10-037-417-46 (1-357) x US-09-654-600A-18 (1-3147)
 QY 32 AlaArgGlyProProTyrCysGlyArgProGluPro----- 43
 Db 1385 AGCAAGGGCAACCTGAGTGTGACGGAAGGAGGACTGTAGCGAGCGCTCAGATGAGAAG 1326
 QY 44 -----SerAlaArgLleValGlyGlySerAsn 52
 Db 1325 GACTGCGACTGGGCTGCGGTCAATTCAGGACAGGCTCGTGTGTTGGGGCAGCGAT 1266
 QY 53 AlaGlnProGlyThrTrrProTrrGlnValSerLeuHisHis---GlyGlyGlyHisIle 71
 Db 1265 GCGATGAGGGCGAGTGGCCCTGGCAGTAAAGCTGCTGCTGCGCAGGCGCCATC 1206
 QY 72 CysGlyGlySerLeuIleAlaProSerTrrValLeuSerAlaAlaHisCysPheMetThr 91
 Db 1205 TGGGTTGCTTCCCTCATCTCTCCAACTGGTGTCTCTGCGCACACATGCTATCATCGAT 1146
 QY 92 AsnGly-----ThrLeuGluProAlaAlaGluTrrSerValLeuLeuGlyValHisSer 109
 Db 1145 GACAGAGGATTCAGGTACTCAGACCCACGCGAGTGGAGCGCCTTCTGGGCTTGCAGAC 1086
 QY 110 Gln---AspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValValPro 128
 Db 1085 CAGAGCCAGGCGAGCGCCCTGGGTTGCAGGAGCGCGAGGCTCAAGCGCATCATCTCCAC 1026
 QY 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuArgLeuAlaSerPro 148
 Db 1025 CCCTTCTCAATGACTTCACTTCGACTATGATGATCGCTGTGGAGCTGGAGAACCG 966
 QY 149 AlaSerLeuGlyProAlaValTrrProValCysLeuProArgAlaSerHisArgPheVal 168
 Db 965 GCAGAGTACAGCTCCATGCTGGGCGCCATCTGCTGCGGAGCGCTCCCATGCTTCCCT 906
 QY 169 HisGlyThrAlaCysTrrAlaThrGlyTrrGlyAspValGlnGluAlaAspProLeuPro 188
 Db 905 CCGGCAAGGCCATCTGGTTCAGCGGCTGGGGACACACCAGTATGGAGGCACTGGGCGCG 846
 QY 189 LeuProTrrValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208
 Db 845 CTG-----ATCCTGCAAAAGGTGAGATCGCGCTCATCAACAGACACCTGGAGAAC 792
 QY 209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCys 228
 Db 791 CTCTGCGCGAG-----CAGATCAGCGCGCATGATGTC 756
 QY 229 AlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeu--- 247
 Db 755 GTGGGCTTCTCAGCGCGCGTGGACTCTCTGCGAGGGTGAATCCGGGGAGCCCTGTGCC 696
 QY 248 ValCysGluGluGlyGlyArgTrrPheGlnAlaGlyTrrSerPheGlyPheGlyCys 267
 Db 695 AGCTGAGGCGGATGGGCGGATCTTCCAGCGCGGTGTGGTGTGGAGTGGGAGACCGGTGC 636
 QY 268 GlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrrPileArg 287
 Db 635 GCTCAGAGGAACAAGCAGCGGTGTACACAGGCTCCCTCTGTTTGGGAGCTGGATCAA 576
 QY 288 GluGlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
 Db 575 GAGAACACTGGGGTATAGGGCGCGGGCCA----- 546
 QY 308 SerAspCysLeuHis 312
 Db 545 CCCAAATGTGTACAC 531

RESULT 13
 US-09-907-794A-262

Sequence 262, Application US/09907794A
Patent No. 6635468
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Pong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavini, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/307,794A
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 262
LENGTH: 1378
TYPE: DNA
ORGANISM: Homo Sapien
US-09-907-794A-262

Alignment Scores:			
Pred. No.:	8-75e-36	Length:	1378
Score:	545.00	Matches:	134
Percent Similarity:	48.00%	Conservative:	46
Best Local Similarity:	35.73%	Mismatches:	146
Query Match:	27.91%	Indels:	51
DB:	4	Gaps:	9
US-10-037-417-46 (1-357) x US-09-907-794A-262 (1-1378)			
QY	7	LeuGlyProGlyGlnLeuGlyAlaValAlaAanSerAspSerTyrSerLeuTyrGlyLeu	26
DB	80	CTGGGTGGGGGCTGTCTCGGCACCTTCACCTCCCTGCTGCTGCTGCTGCGCAGCAATC	139
QY	27	ValProSerGlyProAlaArgGlyProProTyrCysGlyArgProGluProSerAlaArg	46
DB	140	CTCATGCGCGCCAGGATACCTGTTCCCGCAGCTGTGGGAAGCCCGCAGCAGCTGAACCG	199
QY	47	IleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis	66
DB	200	GTTGTGGGGCGGAGCAGCAGCACTGACAGCGAGTGGCCTGGATCGTGCATCCAGAAG	259
QY	67	GlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAla	86
DB	260	AATGGGACCCACCACTGGCGAGGTCTCTGCTCACCAGCGCTGGGTGATCATCTGCTGCC	319
QY	87	HisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGly	106
DB	320	CATGTGTTCAAGACCAAC-----CTGAACAAACCATACCTGTTCTCTGTGCTGCTGGG	373
QY	107	ValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleVal	126
DB	374	GCCTGGCAGCTGGGGAACTCTGCTCGCTCCAGAAAGGTGGGTGTGCTGGGTGGAG	433
QY	127	ValProAlaAsnTyrSerGlnValGlu---LeuGlyAlaAspLeuAlaLeuLeuArgLeu	145
DB	434	CCCCACCTGTGTATCTCTGGAAGAGAGGTGCTGTGCAGACATGCGCTGGTGGCTC	493
QY	146	AlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHis	165
DB	494	GAGCGCTCCATACAGTTCTCAGAGCGGGTCTCGCCCATCTGCTACCTGATCGCTCTATC	553
QY	166	ArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAsp	185
DB	554	CACCTCCCTCCAAACACCCACTGCTGTGATCTCAGGCTGGGGGAGCATCCAGATGGAGTT	613
QY	186	ProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThr	205
DB	614	CCCTTGCCCCACCTCAGACCTCAGAGAGTGAAGGTTCCTATCATCGATCGGAAGTC	673
QY	206	CysGlnCysLeuTyr-----SerGlnProGlyProPheAsnLeuThrLeuGlnIle	222
DB	674	TGACGCCATCTGTACTTGGCGGGAGCAGCAGGAGCCC-----ATC	715
QY	223	LeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAsp	242
DB	716	ACTGAGGACATGCTGTGCGCGGCTACTTGGAGGGGAGCGGGATGCTGTCTGGCGCAC	775
QY	243	SerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSer	262
DB	776	TCCGGGGGCCCCCTCATGTGCAGGTGGACGCGCGCTGGCTGCTGGCGGCATCATCAGC	835
QY	263	PheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyr	282
DB	836	TGGGGCGAGGGCTGTGCGCGAGCGCAACAGCGCGGGGTCTATCATCAGCTCTCTCGGCAC	895
QY	283	GluAlaTrpIleArgGluGln-----	289
DB	896	CGCTCCTGGTGA-GAAGATGTCGAAGGGGTGAGCTCCGCGGGCGCTCAGGGGGG	954
QY	290	-----ValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln	307

Db 955 TGGGGCCCTCAGGACCGACCGAGCGCT-CTGGGGCGCGCGCGCTCTAGGCGCGAG 1013
 QY 308 SerAspCysLeuHisGlnThrAlaPheLeuAspSerAlaArgLeuLeuArgProLeu 327
 Db 1014 CGGAC-----CGGGGCTCGGATCTGAAGG-----CGGCAGAT 1049
 QY 328 SerHisIleSerVal-----GlyValSerThrGlyThrIlyssSerLeu 341
 Db 1050 CCACATCTGGATCTGGATCTGGCGGCTCTGGCGGTTTCCCGCGCGTAATAGGCTC 1109
 QY 342 ValLeuProTrpIleuSerProHisSerLeuLeuGlyLeuTrpGly 356
 Db 1110 ATC-----TACCTTACCTCTGGGG 1130

RESULT 14

US-09-905-125A-262
 ; Sequence 262, Application US/09905125A
 ; Patent No. 6664376
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Baton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, A.
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Hillan, Kenneth, J.
 ; APPLICANT: Kijavlin, Ivar J.
 ; APPLICANT: Mather, Jennie P.
 ; APPLICANT: Pan, James
 ; APPLICANT: Paoni, Nicholas F.
 ; APPLICANT: Roy, Margaret Ann
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Williams, P. Mickey
 ; APPLICANT: Wood, William, I.
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 ; FILE REFERENCE: 10466-14
 ; CURRENT APPLICATION NUMBER: US/09/905,125A
 ; CURRENT FILING DATE: 2001-07-12
 ; PRIOR APPLICATION NUMBER: PCT/US00/04414
 ; PRIOR FILING DATE: 2000-02-22
 ; PRIOR APPLICATION NUMBER: US 60/143,048
 ; PRIOR FILING DATE: 1999-07-07
 ; PRIOR APPLICATION NUMBER: US 60/145,698
 ; PRIOR FILING DATE: 1999-07-26
 ; PRIOR APPLICATION NUMBER: US 60/146,222
 ; PRIOR FILING DATE: 1998-07-28
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594
 ; PRIOR FILING DATE: 1999-09-08
 ; PRIOR APPLICATION NUMBER: PCT/US99/20944
 ; PRIOR FILING DATE: 1999-09-13
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/21547
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/23089
 ; PRIOR FILING DATE: 1999-10-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/28214
 ; PRIOR FILING DATE: 1999-11-29
 ; PRIOR APPLICATION NUMBER: PCT/US99/28313
 ; PRIOR FILING DATE: 1999-11-30
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564
 ; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/28565
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219
 ; PRIOR FILING DATE: 2000-01-05
 ; NUMBER OF SEQ ID NOS: 423
 ; SEQ ID NO 262
 ; LENGTH: 1378
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 ; US-09-905-125A-262
 Alignment Scores:
 Pred. No.: 8,75e-36 Length: 1378
 Score: 545.00 Matches: 134
 Percent Similarity: 48.00% Conservative: 46
 Best Local Similarity: 35.73% Mismatches: 146
 Query Match: 27.91% Indels: 51
 DB: 4 Gaps: 9
 US-10-037-417-46 (1-357) x US-09-905-125A-262 (1-1378)
 QY 7 LeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSerTyrSerLeuTyrGlyLeu 26
 Db 80 CTGGGTGGGGCTGTCTCGGCACCTTCACCTCCCTGCTGCTGCGTCGACAGCCATC 139
 QY 27 ValProSerGlyProAlaArgGlyProProTyrCysGlyArgProGluProSerAlaArg 46
 Db 140 CTCATGCGCGCCAGGATACCTGTTCCTCCCGACCTCTGGGGAAGCCCGACGAGCTGAACCG 199
 QY 47 IleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis 66
 Db 200 GTTGTGGCGCGGAGGACGACGACTGACGCGAGTGGCCCTGGATCGTGAGCATCCAGAAG 259
 QY 67 GlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAla 86
 Db 260 AATGGGACCCACCACTCGCGAGGTTCTCTGCTCACCAGCCGCTGGGTGATCCTCTGCTGCC 319
 QY 87 HisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGly 106
 Db 320 CACTGTTTCAAGGACAAAC-----CTGAACAAACCACTACCTGTCTCTGTGCTGCTGCGG 373
 QY 107 ValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleVal 126
 Db 374 GCCTGGCAGCTGGGGAACCTTGGCTCTCGGTCCAGAGGTGGGTGTTGCTGGGTGGAG 433
 QY 127 ValProAlaAsnTyrSerGlnValGlu-----LeuGlyAlaAspLeuAlaLeuLeuArgLeu 145
 Db 434 CCCACCTCTGTGTATTCTCGAAGGAAGTTCCTGTCAGAGGTCCTGTCAGACATTCCTGCTGCTC 493
 QY 146 AlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHis 165
 Db 494 GAGCGCTCCATACAGTTCTTCAGAGCGGTCCTCGCCATCTGCTGCTGCTGCTGCTGCTATC 553
 QY 166 ArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGlnAlaAsp 185
 Db 554 CACTCCCTCCAAACACCACTGCTGGATCTCAGGTCGCGGAGCATCAAGATGGAGTT 613
 QY 186 ProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThr 205
 Db 614 CCTTGGCCCTCAGACCTGTCAGAGGTCGAGGTTCTCTATCATCATCATCATCATCATCATCAT 673
 QY 206 CysGlnCysLeuTyr-----SerGlnProGlyProPheAsnLeuThrLeuGlnIle 222
 Db 674 TGCAGCCCATCTGTACTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 715
 QY 223 LeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAsp 242

Db 494 GAGCGCTCCATACAGTTCTTCAGAGCGGGTCTCCCATCTGCTACCTGCTGATGCCTCTATC 553
Qy 166 ArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaasp 185
Db 554 CACCTCCCTCCAAACACCCACTGCTGGATCTCAGGCTGGGGAGCATCCAGATGGAGTT 613
Qy 186 ProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeuLeuGlyGluAlaThr 205
Db 614 CCCTTCCCCACCCTCAGACCCCTGCAGAGCTGAGGTTCTTATCATCGACTCGGAGTC 673
Qy 206 CysGlnCysLeuTyr-----SerGlnProGlyProPheAsnLeuThrLeuGlnIle 222
Db 674 TGCAGCCAATCTGTACTGGCGGGGAGCAGGACAGGCC-----ATC 715
Qy 223 LeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGlnGlyAsp 242
Db 716 ACTGAGGACATGCTGTGTCGGCTACTTGGAGGGGAGCGGATGCTTGTCTGGGCGAC 775
Qy 243 SerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSer 262
Db 776 TCCGGGGGCCCTCAGTCCAGGTGGACGGCGCTGGCTGGTGGCCGSCATCATCAGC 835
Qy 263 PheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAlaThrTyr 282
Db 836 TGGGGCGAGGCTGTGCCGAGCGCAACAGGCCCGGGTCTACATCAGCCTCTCTGGGCAC 895
Qy 283 GluAlaTrpIleArgGluGln-----SerGlnProGlyProGlyValPheThrAlaValAlaThrTyr 289
Db 896 CGCTCTCTGGGTGGA-GAAGATCTGTCAAGGGGTGCAGCTCCGCGGGCGCGCTCAGGGGGG 954
Qy 290 -----ValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGln 307
Db 955 TGGGGCCCTCAGGACACCGAGCCAGGGCT-CTGGGGCGCGCGCGCTCCTAGGGCGCAG 1013
Qy 308 SerAspCysLeuHisGlnThrAlaPheLeuAspSerAlaArgIleLeuLeuArgProLeu 327
Db 1014 CGGGAC-----GCGGGCTCGGATCTGAAAGG-----CGGCCAGAT 1049
Qy 328 SerHisIleSerVal-----GlyValSerThrGlyThrLysSerLeu 341
Db 1050 CCACATCTGGATCTGGATCTGGCGGGCTCGGGCGGTTTCCCCCGCCGTAATAGGCTC 1109
Qy 342 ValLeuProTrpLeuSerProHisSerLeuGlyLeuTrpGly 356
Db 1110 ATC-----TACCTCTACCTCTGGGGG 1130

Search completed: February 25, 2004, 18:03:02
Job time : 110 secs

Thu Feb 26 13:50:27 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 17:20:05 ; Search time 399 Seconds
(without alignments)

3133.050 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 1953

Sequence: 1 MAQGVLPQGLGAVANSDS.....TKSLVLPWLSPHSLGLNGF 357

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2308694 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL=frame.p2n.model -DEV=xlp

-Q=/cgn2_1/USPTO.spool_p/US10037417/runat_24022004_101040_17563/app_query.fasta_1.519

-DB=Published Applications NA -QFMT=fastp -SUFFIX=p2n.rnpb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100

-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US10037417@cgn 1 107 @runat 24022004 101040_17563

-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100

-LONGLOG -DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:

2: /cgn2_6/ptodata/1/pubna/PCT_NEW_PUB.seq:

3: /cgn2_6/ptodata/1/pubna/US06_NEW_PUB.seq:

4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq:

5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq:

6: /cgn2_6/ptodata/1/pubna/PCTUS_PUBCOMB.seq:

7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq:

8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq:

9: /cgn2_6/ptodata/1/pubna/US09A_PUBCOMB.seq:

10: /cgn2_6/ptodata/1/pubna/US09B_PUBCOMB.seq:

11: /cgn2_6/ptodata/1/pubna/US09C_PUBCOMB.seq:

12: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq:

13: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq:

14: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq:

15: /cgn2_6/ptodata/1/pubna/US10C_PUBCOMB.seq:

16: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq:

17: /cgn2_6/ptodata/1/pubna/US60_NEW_PUB.seq:

18: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	1505	77.1	2457	9	US-09-888-615-52	Sequence 52, Appl
2	697	35.7	1733	14	US-10-176-847-85	Sequence 85, Appl
3	697	35.7	1834	9	US-09-948-094-1	Sequence 1, Appl
4	697	35.7	1834	9	US-09-880-107-2214	Sequence 2214, Ap
5	697	35.7	1834	9	US-09-967-768A-141	Sequence 141, App
6	697	35.7	1834	14	US-10-097-340-261	Sequence 261, App
7	697	35.7	3382	14	US-10-101-510-447	Sequence 447, App
8	665	34.1	1668	9	US-09-925-301-208	Sequence 208, App
9	627.5	32.1	1726	12	US-10-042-865-29	Sequence 29, Appl
10	612	31.3	1797	14	US-10-109-616-1	Sequence 1, Appl
11	603.5	30.9	1613	13	US-10-042-865-31	Sequence 31, Appl
12	603.5	30.9	1613	13	US-10-041-400A-1	Sequence 1, Appl
13	603.5	30.9	1613	13	US-10-041-264A-1	Sequence 1, Appl
14	591	30.3	1020	15	US-10-042-091A-1	Sequence 25, Appl
15	582.5	29.8	1958	16	US-10-051-874-25	Sequence 29, Appl
16	575.5	29.5	768	14	US-10-221-097-10	Sequence 10, Appl
17	575.5	29.5	882	12	US-10-042-865-33	Sequence 33, Appl
18	571.5	29.3	882	12	US-10-042-865-34	Sequence 34, Appl
19	565	28.9	1887	9	US-09-888-615-27	Sequence 27, Appl
20	565	28.9	1973	14	US-10-190-030B-15	Sequence 15, Appl
21	565	28.9	1110	13	US-10-040-655-1	Sequence 1, Appl
22	563.5	28.9	1110	13	US-10-041-006A-1	Sequence 21, App
23	563.5	28.9	1129	12	US-10-147-493-221	Sequence 221, App
24	563.5	28.9	1129	14	US-10-028-072-221	Sequence 221, App
25	563.5	28.9	1129	14	US-10-121-049-221	Sequence 221, App
26	563.5	28.9	1129	14	US-10-123-904-221	Sequence 221, App
27	563.5	28.9	1129	14	US-10-140-470-221	Sequence 221, App
28	563.5	28.9	1129	14	US-10-175-746-221	Sequence 221, App
29	563.5	28.9	1129	14	US-10-176-918-221	Sequence 221, App
30	563.5	28.9	1129	14	US-10-137-865-221	Sequence 221, App
31	563.5	28.9	1129	14	US-10-176-921-221	Sequence 221, App
32	563.5	28.9	1129	14	US-10-140-474-221	Sequence 221, App
33	563.5	28.9	1129	14	US-10-142-431-221	Sequence 221, App
34	563.5	28.9	1129	14	US-10-143-114-221	Sequence 221, App
35	563.5	28.9	1129	14	US-10-140-002-221	Sequence 221, App
36	563.5	28.9	1129	14	US-10-142-419-221	Sequence 221, App
37	563.5	28.9	1129	14	US-10-123-262-221	Sequence 221, App
38	563.5	28.9	1129	14	US-10-142-423-221	Sequence 221, App
39	563.5	28.9	1129	14	US-10-121-050-221	Sequence 221, App
40	563.5	28.9	1129	14	US-10-141-755-221	Sequence 221, App
41	563.5	28.9	1129	14	US-10-143-032-221	Sequence 221, App
42	563.5	28.9	1129	14	US-10-123-108-221	Sequence 221, App
43	563.5	28.9	1129	14	US-10-123-236-221	Sequence 221, App
44	563.5	28.9	1129	14	US-10-123-261-221	Sequence 221, App
45	563.5	28.9	1129	14	US-10-123-261-221	Sequence 221, App

ALIGNMENTS

RESULT 1
US-09-888-615-52
; Sequence 52, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHITE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUPARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: 2001-06-26
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-888-615-52

; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
 ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
 ; FILE REFERENCE: WEI-039
 ; CURRENT APPLICATION NUMBER: US/10/176,847
 ; NUMBER OF SEQ ID NOS: 112
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 85
 ; LENGTH: 1733
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-176-847-85

Alignment Scores:
 Pred. No.: 1.07e-126 Length: 2457
 Score: 1505.00 Matches: 272
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 77.06% Indels: 0
 DB: Gaps: 9

US-10-037-417-46 (1-357) x US-09-888-615-52 (1-2457)

QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
 DB 112 TCGGGGGCGCCCTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAACCGCGAGCGGGCACC 171
 QY 58 TrpProTyrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIle 77
 DB 172 TGGCCTTGGCAAGTGAAGCTGACCATGGAGTGGCCACATCTGGGGGGGCTCCCTTCATC 231
 QY 78 AlaProSerTyrValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
 DB 232 GCGCCCTCTCGGCTCTCTCCGCTGCTCACTGTCTTTCATGACGANTGGAGCGCTGGAGCCC 291
 QY 98 AlaAlaGluTyrSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
 DB 292 GCGGCGGAGTGTGCTGCTGCGGCGTGCATCTCCAGCAGCGGGCCCTGGAGCGCGCG 351
 QY 118 HisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGluLeuGly 137
 DB 352 CACACCGCGCGATGCGGCCCATCTGTTGGTGGCGGCCAATAAGCAAGTGGAGCTGGGC 411
 QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyrPro 157
 DB 412 GCGGACCTGGCCCTGCTGGCTGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 471
 QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
 DB 472 GATCTGCTGCGCCCGCGCTCACACCGCTTCGTGACCGGACCGCTGCTGGGCGCACCGGC 531
 QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGluValGlu 197
 DB 532 TGGGAGAGAGCTCCAGAGGAGAGATCTCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCT 591
 QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
 DB 592 CTAAGGCTGCTGGCGAGGCGCCACTGTCATATGTCCTACAGCCAGCGCCGCTCCCTCAAC 651
 QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGlnGlyArgArgAsp 237
 DB 652 CTCACCTCCAGATATTGCCAGGAGTGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 711
 QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGlyGlyArgTyrPheGln 257
 DB 712 ACCTGCGAGGAGTGAATCTGGGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 771
 QY 258 AlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
 DB 772 GCAGGATACACAGCTTTGGCTTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTT 831
 QY 278 AlaValAlaThrTyrGluAlaTyrPheArgGluGlnValMetGlySerGluProGlyPro 297
 DB 832 GCTGTGGCTACCTATGAGGATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 891
 QY 298 AlaPheProThrGlnProGlnValThrGlnSerAsp 309
 DB 892 GCCTTTTCCACCCAGCCCGCCAGAGACCCAGTCAGAT 927

RESULT 2

US-10-176-847-85
 ; Sequence 85, Application US/10176847
 ; Publication No. US20030068636A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Veibay, Petter Ole

Alignment Scores:
 Pred. No.: 8.57e-54 Length: 1733
 Score: 697.00 Matches: 160
 Percent Similarity: 55.29% Conservativity: 49
 Best Local Similarity: 42.33% Mismatches: 133
 Query Match: 35.69% Indels: 36
 DB: Gaps: 11

US-10-037-417-46 (1-357) x US-10-176-847-85 (1-1733)

QY 1 MetAlaGlnIysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
 DB 99 ATGGCCCAAGAGGGGGTCTGGGGGCTGGGAGCTGGGGGGCTGGGCC----- 146
 QY 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
 DB 147 ATTCTGCTCTATCTTGGATTACTCCGGTCAGGACAGGAGCGAAGGGCAGAAGCTCCC 206
 QY 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
 DB 207 -----TGGCGT---GTGGCGCCCAAGCAGCATCACAGGTGGCAGGTGAGTGGCC 257
 QY 56 GlyThrTyrProTyrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
 DB 258 GGTCTAGTGGCCCTGGCAGGTGAGCATCACCTATGAGGGGCTCCATGTGTGTGTGTCT 317
 QY 76 LeuIleAlaProSerTyrValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
 DB 318 CTCGCTGTGAGCAGTGGGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 377
 QY 96 GluProAlaAlaGluTyrSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
 DB 378 GAA-----GCCTATGAGGTCAAGCTGGGGGGCCCAACAGCTAGACTTCTTCCGAG 428
 QY 116 GlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGlu 135
 DB 429 GAGCCCAAGGTGAGCAGCCCTGAGGAGCATCATCCCCCCCCCAGCTACCTCCAGGAGGC 488
 QY 136 LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
 DB 489 TCCCAAGGCGGACATTGCACTCTCCCACTCAGCAGACCCATCACCTTCTCCGCTATC 548
 QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
 DB 549 CGGCCCATCTGCTCTCCCTGCGAGCCAAACGCTCTCTCCCAACGCGCTCCACTCCACTG 608
 QY 176 ThrGlyTyrGlyAspValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGlu 195
 DB 609 ACTGGCTGGGGTCTATGTGGCCCCCTCAGTGAGCTCTTGACGCGCCCAAGCCACTTCA 668
 QY 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
 DB 669 CTCGAGTGGCTCTGATCAGTCGTGAGACGTGTAACTGCTGCTGTACACATCGACGCGAA 728
 QY 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
 DB 729 CCTGAGGAGCGGCATTTGTCCCAAGGAGCATGTGTGTGTGTGTGTGTGTGTGTGTGT 788
 QY 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGlyGlyArgTyr 255

```

Db      789  AAGGAGCCTGCAGAGTGAAGTCTGGGGCCCACTCTCCTGCCTGTGGAGGCTCTCTGG 848
Qy      256  PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db      849  TACCTGACGGGCACTTGTAGCTGGGAGATGCTGTGGGGCCCGCAACACAGGCTGTGTG 908
Qy      276  PheThrAlaValAlaThrTyrGluAlaThrTyrGluAlaThrTyrGluAlaMetGlySerGluPro 295
Db      909  TACACTCTGGCTCCTGAGCTATGCTGCTGATCCAAAGCAAGTG-----ACAGAACTC 962
Qy      296  GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
Db      963  CAGCTCTGTGTGGTGGCCCAACCCAGAGAGTCCAGCCCGACAGCAACCTCTGTGGCAGC 1022
Qy      313  GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db      1023  CACCTGGCTTCAGCTCTGCCCCAGCCAGGGCTTGTGAGGGCCCATCTTTCTGCCT 1082
Qy      332  ValGlyValSerThrGlyThrLysSerLeuValLeuProThrLeuSer----- 347
Db      1083  CTGGGCTGGCTGGGC-----CTCCTCTCCCATGGCTCAGCGAGCACTGAGCT 1133
Qy      348  -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db      1134  GGCCCTACTTCCAGATGGATGCATCATCAAGGACAGGAGCTGTGCTCTTC 1187

RESULT 3
US-09-948-094-1
; Sequence 1, Application US/09948094
; Patent No. US2002090625A1
; GENERAL INFORMATION:
; APPLICANT: Mok, Samuel
; APPLICANT: Wong, Kwong-kwok
; TITLE OF INVENTION: Methods of Detecting Cancer Based on Prostate
; FILE REFERENCE: 81994/282423
; CURRENT APPLICATION NUMBER: US/09/948,094
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (229)..(1260)
US-09-948-094-1

Alignment Scores:
Pred. No.: 9,118-54 Length: 1834
Score: 697.00 Matches: 160
Percent Similarity: 55.29% Conservative: 49
Best Local Similarity: 42.33% Mismatches: 133
Query Match: 35.69% Indels: 36
DB: 9 Gaps: 11

US-10-037-417-46 (1-357) x US-09-948-094-1 (1-1834)
Qy      1  MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuLeuAlaValAlaAsnSerAspSer 20
Db      229  ATGGCCCAAGAGGGGTCTCTGGGCTTGGGAGCTGGGGGCTGTGGCC----- 276
Qy      21  TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db      277  ATTCTGCTCTATCTTGGATTACTCTCCGTCGGGACAGGAGCGGAGGGGCAAGCTCCC 336
Qy      36  ProTyrCysGlyArgProGluProSerAlaArgIleValGlySerAsnAlaGlnPro 55
Db      337  -----TGCGGT---GTGGCCCCCAAGCACGCATCACAGGTGGCAGCAGTGCAGTCGCC 387
Qy      56  GlyThrTrpProThrGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySer 75

```

```

Db      388  GGTCAAGTGGCCCTGGCAGGTGAGTATGATGAGCGGTCCATGTGTGTGGTGGCTCT 447
Qy      76  LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db      448  CTCGTGTCTGAGCAGTGGTGTCTGTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 507
Qy      96  GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db      508  GAA-----GCCTATGAGGTCAAGTGGGGGCCACAGCTAGACTCTCTATCTCCGAG 558
Qy      116  GlyAlaHisThrArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGlu 135
Db      559  GAGCCCAAGGTGAGCAGCAGTCAATCAATCAATCAATCAATCAATCAATCAATCAAT 618
Qy      136  LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db      619  TCCAGGGGCGACATGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 678
Qy      156  TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db      679  CGGCCCATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 738
Qy      176  ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
Db      739  ACTGGCTGGGGTCAATGTGGCCCTCTGAGGAGCTCTCTGAGCCCAAGCCACTGAGCAA 798
Qy      196  ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db      799  CTGAGGTGCTCTGATGAGTGTGAGTGTAACTGCTGTACACATCGACGCGCAAG 858
Qy      216  PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235
Db      859  CCTGAGGAGCGCCTTGTCTGAGGAGGACATGCTGTGTCTGTCTGTGTGTGTGTGTG 918
Qy      236  ArgAspThrCysGlnGlyAspSerGlyProLeuValCysGluGluGlyGlyArgTrp 255
Db      919  AAGGACGCTGCGCAGGGTGACTCTGGGGGCGCCACTCTCTGCTCTGTGGAGGGTCTCTGG 978
Qy      256  PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db      979  TACCTGACGGGCACTGTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGGCTGTGTGTG 1038
Qy      276  PheThrAlaValAlaThrTyrGluAlaThrTyrGluAlaThrTyrGluAlaMetGlySerGluPro 295
Db      1039  TACACTCTGGCTCCTGAGTATGCTCTCTGATCCAAAGCAAGGTTG-----ACAGAACTC 1092
Qy      296  GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
Db      1093  CAGCTCTGTGTGGTGGCCCAACCCAGAGGTCCAGCCCGACAGCAACCTCTGTGGCAGC 1152
Qy      313  GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db      1153  CACCTGGGCTTCACTCTGCCCCAGCCAGGGCTTGTGTGAGGCCCATCTCTTTCTGTGCT 1212
Qy      332  ValGlyValSerThrGlyThrLysSerLeuValLeuProThrLeuSer----- 347
Db      1213  CTGGGCTGTGGCTGTGGG-----CTCCTCTCTCCCATGGCTCAGCGAGCACTGAGCT 1263
Qy      348  -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db      1264  GGCCCTACTTCCAGATGGATGCATCATCAAGGACAGGAGCTGTGCTCTTC 1317

RESULT 4
US-09-880-107-2214
; Sequence 2214, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

```

US-10-037-417-46 (1-357) x US-09-880-107-2214 (1-1834)

QY 1 MetAlaGlnIysGlyValIleuGlyProGlyGlnIleuGlyAlaValaAlaAsnSerAspSer 20
Dbb 229 ATGCCCCAGAGGGGGCTCTGGGGCCCTGGGCACGTGGGGCTGTGGCC----- 276
QY 21 TyrSerLeuTyr--GlyLeuValProSerGlyPro-----AlaargGlyPro 35
Dbb 277 ATTCTGCTCATCTTGGATTACTCCGCTCGGCACAGGAGCGAAGGGGACAGAGCTCCC 336
QY 36 ProTyrCysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
Dbb 337 -----TGCGGT--GTGGCCCCCAAGCAGCATCACAGTGGCAGCAGTGCAGTCGCC 387
QY 56 GlyThrTrpProTglnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySer 75
Dbb 388 GGTCAAGTGGCCCTGGCAGGTAGCATCACTATGAGGGCTCCATGTTGTGGTGGCTCT 447
QY 76 LeuIleAlaProSerTrpValIleuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Dbb 448 CTCGTCTCTGACAGTGGTGTGTGTCAGTGTCTACTGTTCCCCAGCGAGCACCACAAG 507
QY 96 GluProAlaAlaGluTrpSerValIleuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Dbb 508 GAA-----GCCTATGAGTCAAGCTGGGGGGCCACCAGCTAGACTCCTACTCCGAG 558
QY 116 GlyAlaHisThrArgAlaValaAlaIleValProAlaAsnTyrSerGlnValGlu 135
Dbb 559 GAGCCCAAGTTCAGCCTCTGAGGACATCATCCCCACCCAGCTACTCTCCAGAGGGC 618
QY 136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Dbb 619 TCCAGGGCGACATTGCATCTCTCCAACTCAGCAGACCCATCACCTTCTCCCGCTACATC 678
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Dbb 679 CGGCCCATCTGCTCCCTCGAGCCACAGCCCTCTTCCCAACGGGCTCCACTGCCTGTC 738
QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValIleuGlnIu 195
Dbb 739 ACTGCTGGGGTCATGTGGCCCCCTCAGTGAGCTCTCTGACGCCCAAGCCTTCGACAA 798
QY 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysIleuTyrSerGlnProGlyPro 215
Dbb 799 CTCGAGGTGCTCTCATCAGTCTGTGAGACGTGTAACTGCCTGTACACATCGACGCCAAG 858
QY 216 PheAsnLeuThrLeuGlnIleuLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArg 235

US-10-037-417-46 (1-357) x US-09-967-768A-141 (1-1834)

Qy	1	Met	Ala	Gln	Iys	Gly	Val	Ileu	Gly	Pro	Gly	Gln	Ileu	Gly	Ala	Val	Ala	Asn	Ser	Asp	Ser	20
Db	229	ATG	CGC	CAG	AAG	GGG	GCT	CTG	GGG	CGT	GCG	AGC	TGG	GGG	GCT	GTG	CC	-----	-----	-----	276	
Qy	21	Tyr	Ser	Ileu	Tyr	---	Gly	Ileu	Val	Pro	Ser	Gly	Pro	-----	-----	-----	-----	-----	-----	-----	35	
Db	277	ATT	CTG	CTC	TAT	CTT	TGA	TTC	TAC	TCC	GGT	CGG	GC	CAC	AGG	ACG	GAG	CGG	GC	AAG	AGT	336

Thu Feb 26 13:50:27 2004

```

QY 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 508 GAA-----GCCATGAGGTCAAGCTGGGGGCCACACAGCTAGACTCTCTCTCGAG 558
QY 116 GlyAlaHisThrArgAlaValAlaAlaValValProAlaHisThrSerGlnValGlu 135
Db 559 GAGCCCAAGGTCAGCAACCTGAGAGGACATATCCCCACCCAGCTACTCTCCAGGAGGC 618
QY 136 LeuGlyAlaAspLeuAlaLeuLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 619 TCCAGAGGCGACATGACTCTCCAACTCAGAGACCCATCAGCTTCTCTCCGCTATCAT 678
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 679 CGGCCCATCTGCTCTCTGAGCAACAGCCCTCTCCCAAGAGGCTCTCCATGACTGCTGTC 738
QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
Db 739 ACTGCTGGGGTCATGTGGCCCTCTCAGTGGAGCTCTCTGAGCCCAAGCCACTGCAGCAA 798
QY 196 ValGlnLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTySerGlnProGlyPro 215
Db 799 CTCGAGGTGCTCTGATCAGTGTGAGAGCTGTAACTGCTGTACACATCGACGCCCAAG 858
QY 216 PheAsnLeuThrLeuGlnLeuLeuProGlyMetLeuCysAlaGlyTrpProGluGlyArg 235
Db 859 CCTGAGGAGCGCACTTTGTCCAAAGAGGACATGGTGTGCTGTGCTGTGAGGGGTCTCTGG 918
QY 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrp 255
Db 919 AAGGAGCGCTGCGAGGAGTACTCTGGGGGCCCACTCTCTGCTGCTGTGAGGGGTCTCTGG 978
QY 256 PheGlnAlaGlyLeuThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db 979 TACTGAGGGGCACTGTGAGCTGGGAGATGCTGTGGGGCCGCAAGAGGCTGTGTGTG 1038
QY 276 PheThrAlaValAlaThrTrpGluAlaTrpIleArgGluGlnValMetGlySerGluPro 295
Db 1039 TACTGAGGGGCACTGTGAGCTGGGAGATGCTGTGGGGCCGCAAGAGGCTGTGTGTG 1092
QY 296 GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
Db 1093 CAGCCTGTGTGGTGGCCCAACCCAGGAGTCCAGCGCCGACCAACCTCTGTGGCAGC 1152
QY 313 GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db 1153 CACCTGGCCTTCAGCTGCGCCCAAGCCAGGAGTGTCTGAGGGCCATCTTTTCTGCT 1212
QY 332 ValGlyValSerThrGlyThrLysSerLeuValLeuProTrpLeuSer----- 347
Db 1213 CTGGGCTGTGTGGTGGG-----CTGCTCTCCCATGCTCAGCGAGCACTGAGCT 1263
QY 348 -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db 1264 GGCCTTACTTCCAGTGGATGCATCAGCTCAAGGAGGAGGCTGTGTCTCTC 1317

RESULT 7
US-10-101-510-447
; Sequence 447, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; PRIOR FILING DATE: 2002-03-20
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 447
; LENGTH: 3382

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-447
Alignment Scores: 1,76e-53 Length: 3382
Pred. No.: 697,00 Matches: 160
Score: 55.29% Conservative: 49
Percent Similarity: 42.33% Mismatches: 133
Best Local Similarity: 35.69% Indels: 36
Query Match: 14 Gaps: 11
DB:
US-10-037-417-46 (1-357) x US-10-101-510-447 (1-3382)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 230 ATGGCCCAAGAGGGGTCTCTGGGGCTTGGGCGCTGGGGGCTGTGGCC----- 277
QY 21 TyrSerLeuTy---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db 278 ATTCTGCTCTATCTTGGANTACTCCGGTCGGGGACAGGAGCGGAAGGGCAGAAAGCTCCC 337
QY 36 ProTyrcysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnPro 55
Db 338 -----TGCAGT---GTGGCCCCCAAGCAGCGCATCACAGGTGGCAGCAGTCGCC 388
QY 56 GlyThrTrpProTrpGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySer 75
Db 389 GGTGAGTGGCCCTGGCAGGTGAGCATCACCTATGAGGGGTCCATGTGTGTGTGGTCTCT 448
QY 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db 449 CTCGTGTCTGACGAGTGGGTGTGTGTCAGTGTCTACTGCTTCCCGAGCAGCAGCACCAAG 508
QY 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 509 GAA-----GCATATGAGGTCAAGCTGGGGGCCCAACACAGCTAGACTCTCTACTCCGAG 559
QY 116 GlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTrpSerGlnValGlu 135
Db 560 GAGCCCAAGGTCAAGCACCCTGAGGAGCATCCCCCCCCAGCTACTCTCCAGGAGGGC 619
QY 136 LeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 620 TCCAGGGCGACATTGCACTCTCTCCAACTCAGCAGACCCATCATCTCTCCCGCTATCAT 679
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 680 CGGCCCATCTGCTCTCTGCGCCCAAGCAGCGCTCTTCCCAACGCGCTCCACTGCAGTGC 739
QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
Db 740 ACTGGCTGGGGTCTATGTGGCCCCCTCAGTGGAGCTCTCTGAGCCCAAGCCACTGCAGCA 799
QY 196 ValGlnLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTySerGlnProGlyPro 215
Db 800 CTCGAGGTGCTCTGATCAGTGTGAGAGCTGTAACTGCCTGTACAACTCAGAGCCCAAG 859
QY 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTrpProGluGlyArg 235
Db 860 CCTGAGGAGCGGCACCTTTGTTCTCCAAAGAGGACATGTGTGTGTGTGTGTGTGTGTGT 919
QY 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrp 255
Db 920 AAGGAGCGCTGCCAGGTGACTCTGGGGGCCCACTCTCTGCTGTGTGGAGGGTCTCTGG 979
QY 256 PheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db 980 TACTGAGGGGCACTTGTGAGCTGGGGAGATGCTCTGTGGGGCCCGCAAGAGGCTGTGTG 1039
QY 276 PheThrAlaValAlaThrTrpGluAlaTrpIleArgGluGlnValMetGlySerGluPro 295
Db 1040 TACTCTTGGCCCTCCAGCTATGCTCTCTGTGATTCAAAGCAGAGGTG-----ACAGA 1093
;

```


APPLICANT: Casman, Stacie J
APPLICANT: Shenoy, Suresh G
APPLICANT: Spytek, Kimberly
APPLICANT: Zhong, Mei
APPLICANT: Gangolli, Beba A
APPLICANT: Burgess, Catherine E
APPLICANT: Patturajan, Meera
APPLICANT: Vernet, Corine A.M
APPLICANT: Taylor, Sarah
APPLICANT: Tchernev, Velizar T
APPLICANT: Miller, Charles E
APPLICANT: Guo, Xiaojia
APPLICANT: Boldog, Ference L
APPLICANT: Grosse, William M
APPLICANT: Alsbrook II, John P
APPLICANT: Gerlach, Valerie L
APPLICANT: Edinger, Shlomit R
APPLICANT: Rothenberg, Mark E
APPLICANT: Ellerman, Karen
APPLICANT: MacDougall, John
APPLICANT: Malyankar, Uriel M
APPLICANT: Millet, Isabelle
APPLICANT: Peyman, John
APPLICANT: Smithson, Glennda
APPLICANT: Gunther, Erik
APPLICANT: Stone, David
TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
TITLE OF INVENTION: Using the Same
FILE REFERENCE: 21402-337
CURRENT APPLICATION NUMBER: US/10/042,865
CURRENT FILING DATE: 2002-05-17
PRIOR APPLICATION NUMBER: 60/260,417
PRIOR FILING DATE: 2001-01-09
PRIOR APPLICATION NUMBER: 60/260,831
PRIOR FILING DATE: 2001-01-10
PRIOR APPLICATION NUMBER: 60/272,338
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/274,876
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/284,704
PRIOR FILING DATE: 2001-04-18
NUMBER OF SEQ ID NOS: 264
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 29
LENGTH: 1726
TYPE: DNA
ORGANISM: Homo sapiens
US-10-042-865-29
Alignment Scores:
Pred. No.: 1e-49 Length: 1726
Score: 652.00 Matches: 152
Percent Similarity: 52.12% Conservative: 45
Best Local Similarity: 40.21% Mismatches: 109
Query Match: 33.38% Indels: 72
Gaps: 12
DBs:
US-10-037-417-46 (1-357) x US-10-042-865-29 (1-1726)
QY 1 MetAlaGlnTyrGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 229 ATGGCCCAAGAGGGGTCTCTGGGGCTTGGGAGCTGGGAGCTGGGGCTGTGGCC-----276
QY 21 TyrSerLeuTyr---GlyLeuValProSerGlyPro-----AlaArgGlyPro 35
Db 277 ATTCTGCTCTATTCTGGATTACTCCGTCGGGACAGGAGCGGAAGGGGAGAGCTCC 336
QY 36 ProTyrCysGlyArgProGluProSerAlaArgTleValGlyGlySerAenAlaGlnPro 55
Db 337 -----TGCAGT---GTGGGCCCCCAAGACCGCATCACAGTGGCAGCTGAGTCGCC 387
QY 56 GlyThrTrpProTrpGlnValSerLeuHisGlyGlyGlyHisGlyGlyHisGlyGlySer 75

Db 388 GGTCACTGGCCCTGGCAGGTGAGCATCATCTATGAGGCGTCCATGTGTGTGGTGGCTCT 447
QY 76 LeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeu 95
Db 448 CTCGTGTCTGAGCAGTGGGTGTGTGTCAGCTGTCTCTCTCTCTCTCTCTCTCTCTCT 489
QY 96 GluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAsp 115
Db 489 -----489
QY 116 GlyAlaHisThrArgAlaValAlaAlaValValProAlaAsnTyrSerGlnValGlu 135
Db 490 -----CCAGCGAGCACCACCAAGGGCTCC 513
QY 136 LeuGlyAlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaVal 155
Db 514 CAGGGC---GACATTGCACCTCTCCCACTCAGCAGACCCACAGCTACTCCCGCTACATC 570
QY 156 TrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAla 175
Db 571 CGGCCCATCTGCTCTCTCTGAGCAAGCCCTCTTCCCAACGGCTCCCTCCCTGACATGTC 630
QY 176 ThrGlyTrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGlu 195
Db 631 ACTGGCTGGGTGTCATGTGGCCCTCTCAGTCAGCTCTGAGCCCAAGCCACTCAGCAA 690
QY 196 ValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyPro 215
Db 691 CTCGAGGTGCTCTGATCAGCTGTCGAGACGTGTAACTGCTGTACACATCAGCAGCAAG 750
QY 216 PheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGlyGlyArg 235
Db 751 CTTGAGGAGCGCCTTTGTCCAGAGGACATGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 810
QY 236 ArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrp 255
Db 811 AAGGAGCGCTGCGCAGGTGACTCTGGGACCCACTCTCTGCTGCTGCTGCTGCTGCTGCT 870
QY 256 PheGlnAlaGlyTleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyVal 275
Db 871 TACCTGACGGGCAATTGTGAGCTGGGGAGATGCTCTGGGGCCCGCAACAGGCTGTGTGTG 930
QY 276 PheThrAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluPro 295
Db 931 TACACTCTGGCTTCCAGCTATGCTCTCTGATCCAAAGCAAGGTG-----ACAGAACTC 984
QY 296 GlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHis 312
Db 985 CAGCTCTGTGTGTGGTGGCCCAACCCAGGATCCAGCCCGCCAGCAACCTCTGTGGCAGC 1044
QY 313 GlnThrAlaPheLeuAspSer---AlaArgIleLeuLeuArgProLeuSerHisIleSer 331
Db 1045 CACCTGCGCTTTCAGCTCTGCGCCAGCCAGGGCTTGTGAGGGCCATCTTTTCTGCTCT 1104
QY 332 ValGlyValSerThrGlyThrLysSerLeuValLeuProTrpLeuSer-----347
Db 1105 CTGGGCTGT 1155
QY 348 -----ProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db 1156 GGCCCTACTTCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1209
RESULT 10
US-10-109-616-1
; Sequence 1, Application US/10109616
; Publication No. US20030167484A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CHANNEL
; FILE OF INVENTION: ACTIVATING PROTEASE 1 (CAP1) GENE DISRUPTIONS
; FILE REFERENCE: R-490
; CURRENT APPLICATION NUMBER: US/10/109,616
; CURRENT FILING DATE: 2002-03-28

; PRIOR APPLICATION NUMBER: US 60/280,509
 ; PRIOR FILING DATE: 2001-03-29
 ; PRIOR APPLICATION NUMBER: US 60/311,055
 ; PRIOR FILING DATE: 2001-08-08
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 1797
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-10-109-616-1

Alignment Scores:
 Pred. No.: 1,72e-47 Length: 1797
 Score: 627.50 Matches: 144
 Percent Similarity: 55.17% Conservatives: 48
 Best Local Similarity: 41.38% Mismatches: 139
 Query Match: 32.13% Indels: 17
 DB: Gaps: 9

US-10-037-417-46 (1-357) x US-10-109-616-1 (1-1797)

QY 1 MetAlaGlnysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAenSerAspSer 20
 Db 146 ATGCCCCAAGGTTGGGCTTGGGACTTGGGAGCTGGAGCTGTGACC-----ATT 196
 QY 21 TyrSerLeuTyrglyLeuValProSerGly---ProAlaargGly---ProProTyrcys 38
 Db 197 CTGCTCTCTCGGATTCCTCCAGTCCGGAATCCGAGTCCGAGTCCGAGCTCCCTGT 256
 QY 39 GlyArgPro---GluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
 Db 257 GGTCCCTTCATCCAGCCA-----CGATCACCGTGTGTGCGAGTGCAGGCCGGTCCAG 310
 QY 58 TrpProTrpGlnValSerLeuHisGlyGlyGlyHisGlyCysGlyGlySerLeuIle 77
 Db 311 TGGCCCTTGGCAGTCCAGTCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 370
 QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
 Db 371 TCAATAAATGGGTGGTGTCTGCTGCTACACGCTTCCCGAGAGAACAGCAGCGGAA--- 427
 QY 98 AlaAlaGluTrpSerValLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
 Db 428 -----GGTATGAGTGAAGTGGGGCCCGCCAGCTAGACTCTCAGCAATGACACT 481
 QY 118 HisThrArgAlaValAlaAlaIleValProAlaAenTyrsGlnValGluLeuGly 137
 Db 482 GTGGTCCACACAGTGGCTCAGATCATCCACCCTCAGCTACCGAGAGAGAGGGTCCCGAG 541
 QY 138 AlaAspLeuAlaLeuLeuArgLeuSerProAlaSerLeuGlyProAlaValTrpPro 157
 Db 542 GGGGACATCGGTTTCATCCGCTAGCAGTCTGTCTCCAGCGGCTTCCCGGTATCATCAGACC 601
 QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
 Db 602 ATTGCTCTCCGCGCAATAGCTCTCTCCAGCGGCTTCCAGCTTCTCCCGGTATCATCAGACC 661
 QY 178 TrpGlyAspValGlnGluAlaAspProLeuProTrpValLeuGlnGluValGlu 197
 Db 662 TGGGGTCATGTGGCTCTTCTAGTGGCTTCCAGAGCTCCAGAGCTTCTCCAGAGCTCCAG 721
 QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrsSerGlnProGlyProPheAen 217
 Db 722 GTACCATCATCAGCGGGAACCTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781
 QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrsProGluGlyArgArgAsp 237
 Db 782 GAACCGCACACTATCCAGCAGGACATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 841
 QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGln 257
 Db 842 GCCTGCCAGGGTGACTCTGGGGGCCCACTCTCTTGTCCAGAGGAGGAGGAGGAGGAGGAG 901

QY 258 AlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
 Db 902 GAGGCAATGTGAGTGGGTGATGCCGTGTGGTGGCCCAACAGGCGCTGGAGTATACACT 961
 QY 278 AlaValAlaThrTyrgluAlaTrpIleArgGluGlnValMetGlySerGluProGlyPro 297
 Db 962 CTGACTTCTACCTATGCTTCTCTGGATCCACCATGTG-----GCAGAGCTCCAGCCT 1015
 QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp-----CysLeuHisGlnThr 314
 Db 1016 CGAGTTGTCCCGCAGACTCCAGAGTCCCGAGTCCCGAGTCCCGAGTCCCGAGTCCCG 1075
 QY 315 AlaPheLeuAspSerAlaArg---IleLeuLeuArgProLeuSerHisSerValGly 333
 Db 1076 GTCTTCAGCTCAGCGGCGAGCCCGAAATTTGTTAAGCCCGGTACTTTTCTGCCACTTGGT 1135
 QY 334 ValSerThrGlyThrLysSerLeu 341
 Db 1136 CTGACCTGGGCTCTTGTCTTTG 1159

RESULT 11

US-10-042-865-31
 ; Sequence 31, Application US/10042865
 ; Publication NO. US20040029216A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Padigar, Muralidhara
 ; APPLICANT: Li, Li
 ; APPLICANT: Zerhusen, Bryan D
 ; APPLICANT: Casman, Stacie J
 ; APPLICANT: Shenoy, Suresh G
 ; APPLICANT: Spytek, Kimberly
 ; APPLICANT: Zhong, Mei
 ; APPLICANT: Gangolli, Esha A
 ; APPLICANT: Burgess, Catherine E
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Vernet, Corine A.M
 ; APPLICANT: Taylor, Sarah
 ; APPLICANT: Tchernev, Velizar T
 ; APPLICANT: Miller, Charles E
 ; APPLICANT: Guo, Xiaojia
 ; APPLICANT: Boldog, Ference L
 ; APPLICANT: Grosse, William M
 ; APPLICANT: Alsbrook II, John P
 ; APPLICANT: Gerlach, Valerie L
 ; APPLICANT: Edinger, Shlomit R
 ; APPLICANT: Rothenberg, Mark E
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: MacDougall, John
 ; APPLICANT: Malyankar, Uriel M
 ; APPLICANT: Millet, Isabelle
 ; APPLICANT: Peyman, John
 ; APPLICANT: Smithson, Glennda
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Stone, David
 ; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
 ; FILE REFERENCE: 21402-537
 ; CURRENT APPLICATION NUMBER: US/10/042,865
 ; CURRENT FILING DATE: 2002-05-17
 ; PRIOR FILING DATE: 2001-01-09
 ; PRIOR APPLICATION NUMBER: 60/260,417
 ; PRIOR FILING DATE: 2001-01-09
 ; PRIOR APPLICATION NUMBER: 60/260,831
 ; PRIOR FILING DATE: 2001-01-10
 ; PRIOR APPLICATION NUMBER: 60/272,338
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: 60/274,876
 ; PRIOR FILING DATE: 2001-03-09
 ; PRIOR APPLICATION NUMBER: 60/284,704
 ; PRIOR FILING DATE: 2001-04-18
 ; NUMBER OF SEQ ID NOS: 264
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 31


```

441  GGCGACCTGGCAGCTGCTGCGTGGCGGGTGGCCCTGAGCGCTGCGGCTCCAAACC 500
158  ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
501  GHTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
178  TrpGlyAspValGlnGluAlaAspProLeuProLeuProLeuProTrpValLeuGlnValGlu 197
561  TGGGGCAGCGCTCCGCCAGAGTGGCCCTCCAGAGTGGCGACCGCTACAAGAGGTAAGG 620
198  LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyr-----SerGlnPro 213
621  TGGCGCTGCTGGACTCGCGCACCTGCGACGGCTCTACACGTGGGCGCGACGTGCC 680
214  GlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGlu 233
681  CAGGCTGAGCGATT-----GTGCTGCCTGGGAGTCTGTGTCCGGTACCCCCAG 731
234  GlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGly 253
732  GGCCACAAGAGCGCTGCCAGGGTGATTCTGGGGGACCTCTGCACCTGCCTGCAGTCTGG 791
254  ArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgPro 273
792  AGCTGGGTCTGTGGTGGCGGTGGTGGTGGCGAAGGGTTGTGCCCTGCCCAACCTCCA 851
274  GlyValPheThrAlaValAlaAlaThrTyrGluAlaTrpIleArgGlnGlnValMetGlySer 293
852  GGGGTCTACACCAAGTGTGCCACATATAGCCCTGGATTCAAGGCTCGCGTCACTTCTAAT 911
294  Glu-----ProGly----- 296
912  GCTAGCGGTGAGGTGACCTGGAGCGAGTGTGGGTGCTCAGCCTCCTGGTTCATC 971
297  -----ProAlaPhe-ProThrGlnProGlnIlyThrGlnSerAspCysLeuHisGlnTh 314
972  CAGGCACCTGCCTATACCCACATCCCTTCTGCTCGAGGCCAAGATGCCTAAAAAG-- 1029
314  rAlaPheLeuAspSerAlaArgIleLeuLeuArgProLeuSerHisIleSerValGlyVa 334
1030  -----CTAAGGCCACCCACCCCGCCAGCCCTTCTGGCTCCTCT 1073
334  lSerThrGlyThylsSerLeuValLeuProTrpLeuSer----- 347
1074  CTTCTTTGGGATCACCAGCTCTGACTTCACCAACCTCTATCCAGGAATCTGCCATGAGT 1133
348  -----ProHisSerLeuLeuGlyLeuTrp 355
1134  CCCAGGAGTCACACTCCCACTCCCTTCTGGCTTGAT 1173

RESULT 13
US-10-041-264A-1
; Sequence 1, Application US/10041264A
; Publication No. US20020142446A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,264A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-041-264A-1

```

Alignment Scores:
 Pred. No.: 2,266-45 Length: 1613
 Score: 603.50 Matches: 139
 Percent Similarity: 50.08% Conservative: 38
 Best Local Similarity: 39.27% Mismatches: 129
 Query Match: 30.90% Indels: 48
 DB: 13 Gaps: 7

US-10-037-417-46 (1-357) x US-10-041-264A-1 (1-1613)

QY	38	CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr	57
DB	150	TGCGGGCAGCCCGCATGTCACAGTCGATCGTTGGGGCGCGGATGGCCGGACGGAG	209
QY	58	TrpProTrpGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeuIle	77
DB	210	TGCGCGTGGCAGCGAGCATCCAGCATCTCTGGGGCACACGTGTGGGGGGTTCGCTCATC	269
QY	78	AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro	97
DB	270	GCCTCCCGAGTGGTGTGACAGCGGCGCATCTGTTCCCGAGGAGGACATG-----	320
QY	98	AlaAlaGlnTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla	117
DB	321	CCAGCTCAGTACCGCGTGGCGCTCTGGGGGGCGTCTGGGCTCCACCTCGCCCGCCACG	380
QY	118	HisThrArgAlaValAlaAlaIleValValProAlaAsnTrpSerGlnValGluLeuGly	137
DB	381	CTCTCGTGGCGGTGGCAGCGGTGTCTGTGCCCGCGACTACTCCGAGGACGGGGCCGC	440
QY	138	AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro	157
DB	441	GGCGACCTGCACCTGTCTGCAGCTCGCTGCGCGCGCGTGCCTCGAGCGCTCGCGTCCA	500
QY	158	ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly	177
DB	501	GTCGTGCTGCCGTGCCGGGGCGCGCCCGCGCGCGGACACCATGCGGGTCAACGGC	560
QY	178	TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu	197
DB	561	TGGGGCAGCTCCGCCCGCAGGAGTGCCCTCCAGAGTGGCGACCGGTACAAAGAGTAAGG	620
QY	198	LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyr-----SerGlnPro	213
DB	621	TGTCCGTGTGTGACTCGCGACCTCGCGACCGCTCTACCACGTGGCGCGGACGTGCC	680
QY	214	GlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGlu	233
DB	681	CAGGCTGAGCGCATY-----GTGCTGCTGGGAGTCTGTGTGCGGCTACCCCGCAG	731
QY	234	GlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGly	253
DB	732	GGCCACAAGCAGCGCTGCGAGGGTGTATCTGGGGGACCTCTGCACCTCGCTGCACTGGG	791
QY	254	ArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyAspAsgAsnArgPro	273
DB	792	AGCTGGTCTCTGGTGGGGGTGGTAGCTGGGGCAGGGTTGTGCCCTGCCCAACCGTCCA	851
QY	274	GlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySer	293
DB	852	GGGGTCTACACAGTGTGGCCACATATAGCCCTTGGATTGAGGCTCGCGTCACTTCTAAT	911
QY	294	Glu-----ProGly-----	296
DB	912	GCTAGCCGGTGAGGTGACCTGGAGCCAGCTGTGGGGTCCCTCAGGCTCTCTGGTTTCATC	971
QY	297	-----ProAlaPhe-ProThrGlnProGlnIleThrGlnSerAspCysLeuHisGlnThr	314
DB	972	CAGGCACCTGTCTATACCCACATCCCTTCTGTCCCTCGAGGCCAAGATGCTTAAAAAG--	1029
QY	314	rAlaPheLeuAspSerAlaArgIleLeuLeuArgProLeuSerHisIleSerValGlyIle	334
DB	1030	-----CTAAAGGCCACCCCAACCCCGCCACCCACCTTGTGGCTCTCTCT	1073

QY 334 lserThrGlyThrIleValLeuProTrpLeuSer----- 347
 Db 1074 CCTTTTGGGGATACACAGCTCTGACCTCCACCCACCTCTCCAGGATTCGCATGAGT 1133
 QY 348 -----ProHisSerLeuLeuGlyLeuTrp 355
 Db 1134 CCCAGGAGTCCACTCCCTTCTCTGGCTGTAT 1173

RESULT 14
 US-10-042-091A-1
 ; Sequence 1, Application US/10042091A
 ; Publication No. US20020142447A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Darrow, Andrew
 ; APPLICANT: Andrade-Gordon, Patricia
 ; APPLICANT: Qi, Jensen
 ; TITLE OF INVENTION: DNA Encoding the Human Serine
 ; TITLE OF INVENTION: Protease EOS
 ; FILE REFERENCE: ORT-1031
 ; CURRENT APPLICATION NUMBER: US/10/042,091A
 ; CURRENT FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US/09/387,375
 ; PRIOR FILING DATE: 1999-08-31
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1613
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-042-091A-1

Alignment Scores:
 Pred. No.: 2,268-45 Length: 1613
 Score: 603.50 Matches: 139
 Percent Similarity: 50.00% Conservatives: 38
 Best Local Similarity: 39.27% Mismatches: 129
 Query Match: 30.90% Indels: 48
 DB: 13 Gaps: 7

US-10-037-417-46 (1-357) x US-10-042-091A-1 (1-1613)

QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
 Db 150 TGGGGCAGCCCGCATGCTCCAGTGGTGGGGCGGGATGGCGGGAGAG 209
 QY 58 TrpProTrpGlnValSerLeuHisGlyGlyGlyHisGlyGlyGlySerLeuLeu 77
 Db 210 TGGCGTGGCAGCGGAGCATCCAGCATCTCTGGGGCACAGTGTGGGGGGTGCCTCATC 269
 QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
 Db 270 GCCCCCCAGTGGGTGTGACAGCGGGCGACTGTTTCCCGAGGAGGCGACTG----- 320
 QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
 Db 321 CCAGCTGAGTACCGCGTGGCTGGGGCGCTCGTCTGGGGCTCCACTCGCCCCCGCAGC 380
 QY 118 HisThrArgAlaValAlaAlaIleValProAlaAsnTrpSerGlnValGluLeuGly 137
 Db 381 CTCTCGTGGTGGCGAGCGGCTGCTGCTGCCCCCGGAGTACTCCGAGGAGCGGGCCCGC 440
 QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
 Db 441 GCGGACTGGGACTGCTGCGAGTGGTGGCGCGCGTGGCTGGCGGCTCGCGTCCAAACCC 500
 QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
 Db 501 GTCTGCTGCG 560
 QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGluValGlu 197
 Db 561 TGGGGCAGCCTCCCGCCCGAGGAGTGGCTCCCGCTCCAGAGTGGCGACCGCTACAGAGTAA 620

QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTrp-----SerGlnPro 213
 Db 621 GTGCGCTGCTGAGTCTGGCGACCTCTGACAGGCTCTACACAGTGGGGCGGGAGCTGCC 680
 QY 214 GlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTrpProGlu 233
 Db 681 CAGGCTGAGCGCAT-----GTGCTGGCTGGGAGTCTGTGTGCGGCTACCCCGCAG 731
 QY 234 GlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGly 253
 Db 732 GGCCACAGGAGCGCTGCCAGGTGATCTGGGGGACCTCTGACCTGCTGCGCTGCGGTGGG 791
 QY 254 ArgTrpPheGlnAlaGlyIleThrSerPheGlyCysGlyArgArgAsnArgPro 273
 Db 792 AGCTGGGTCTCTGGTGGCGTGGTGGAGTGGGGCAAGGGTGTGGCTGCTGCTGCTCTCA 851
 QY 274 GlyValPheThrAlaValAlaThrTrpGluAlaTrpIleArgGluGlnValMetGlySer 293
 Db 852 GGGGTCTACACAGTGTGGGCACATATAGCCCTGGATTCAGGCTGCGCTCACTTCTAAT 911
 QY 294 Glu-----ProGly----- 296
 Db 912 GCTAGCGGTGAGGCTGACCTGGAGCCAGCTGCTGGGGTCCCTCAGCTCTGCTGTTTCATC 971
 QY 297 -----ProAlaPhe-ProThrGlnProGlnLysThrGlnSerAspCysLeuHisGlnTh 314
 Db 972 CAGGACCTGCTATACCCACATCCCTTCTGCTCGAGCCCAAGATGCTTAAAAAG-- 1029
 QY 314 rAlaPheLeuAspSerAlaArgIleLeuLeuArgProLeuSerHisLeuValGlyVa 334
 Db 1030 -----CTAAAGGCCACCCACCCACCCACCCACCTTCTGGCTCTCT 1073
 QY 334 lserThrGlyThrLysSerLeuValLeuProTrpLeuSer----- 347
 Db 1074 CCTTTTGGGGATCACAGCTCTGACTCCACACCTCTCCAGGATTCGCCATGAGT 1133
 QY 348 -----ProHisSerLeuLeuGlyLeuTrp 355
 Db 1134 CCCAGGAGTCCACTCCCTTCTCTGGCTGTAT 1173

RESULT 15
 US-10-051-874-25
 ; Sequence 25, Application US/10051874
 ; Publication No. US2004000557A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Alsobrook II, John P
 ; APPLICANT: Colman, Steven D
 ; APPLICANT: Spytek, Kimberly A
 ; APPLICANT: Boldog, Ferenc
 ; APPLICANT: Vernet, Corine AM
 ; APPLICANT: Li, Li
 ; APPLICANT: Shenoy, Suresh G
 ; APPLICANT: Casman, Stacie J
 ; APPLICANT: Guo, Xiaojia Sasha
 ; APPLICANT: Edinger, Shlomit R
 ; APPLICANT: MacDougall, John R
 ; APPLICANT: Malyankar, Uriel M
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Shimkets, Richard A
 ; APPLICANT: Pena, Carol EA
 ; APPLICANT: Tchernev, Velizar T
 ; APPLICANT: Zernusen, Bryan D
 ; APPLICANT: Millet, Isabelle
 ; APPLICANT: Miller, Charles E
 ; APPLICANT: Lepley, Denise M
 ; APPLICANT: Smithson, Glennda
 ; APPLICANT: Baumgartner, Jason C
 ; APPLICANT: Herrman, John L
 ; APPLICANT: Peyman, John A
 ; APPLICANT: Gorman, Linda
 ; APPLICANT: Mezes, Peter D

Thu Feb 26 13:50:27 2004

APPLICANT: Kekuda, Ramesh
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Gerlach, Valerie
APPLICANT: Grosse, William M
APPLICANT: Liu, Xiaohong
APPLICANT: Ellerman, Karen
APPLICANT: Rothenberg, Mark
APPLICANT: Stone, David J
APPLICANT: Burgess, Catherine E
TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
FILE REFERENCE: 21402-245
CURRENT APPLICATION NUMBER: US/10/051,874
CURRENT FILING DATE: 2002-09-25
PRIOR APPLICATION NUMBER: 60/268,595
PRIOR FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: 60/325,306
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/262,587
PRIOR FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: 60/272,409
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/262,454
PRIOR FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: 60/276,777
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/291,672
PRIOR FILING DATE: 2001-05-17
PRIOR APPLICATION NUMBER: 60/330,336
PRIOR FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: 60/265,530
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/261,376
PRIOR FILING DATE: 2001-01-16
NUMBER OF SEQ ID NOS: 269
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 25
LENGTH: 1020
TYPE: DNA
ORGANISM: Homo sapiens
US-10-051-874-25

Alignment Scores:
Pred. No.: 1,866-44 Length: 1020
Score: 591.00 Matches: 125
Percent Similarity: 54.17% Conservative: 31
Best Local Similarity: 43.40% Mismatches: 102
Query Match: 30.26% Indels: 30
DB: 15 Gaps: 6

US-10-037-417-46 (1-357) x US-10-051-874-25 (1-1020)

QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
Db 172 TGGGGGAGCCCGCATGTCAGTTCGATCGTTGGGGCCCGGGATGCCGGGAGAG 231
QY 58 TrpProTrpGlnValSerLeuHisHisGlyGlyHisIleCysGlyGlySerLeuIle 77
Db 232 TGGCCGTGGGAGCGGAGCATTCAGTCGTGGGGGACACCGTGTGGGGGGTGGCTATC 291
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
Db 292 GCCCCCGAGTGGTGGTGGAGCGGCGGACCTGCTCCCGAGGGGACATG----- 342
QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
Db 343 CCAGCTGAGTACCGCGTGGCGCTCGCTGGGGCGGCTGGCTCCCGCTCCACCTGCCCGGCGAG 402
QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAntySerGlnValGluLeuGly 137
Db 403 CTCCTGGTGGCGGAGGGTGTGCTGCTCCCGGAGCTACTTCGAGGAGGGGCGCGC 462
QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157

Db 463 GGGACCTGGACATGCTGCAGCTGCGTGGCGGCTGCCCTGAGCGTCCGCTCCACCC 522
QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
Db 523 GTCTGCTGCTGGCGTGGCGGCGCGCGCGCGCGCACACCATGCGGGTCCAGGCG 582
QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
Db 583 TGGGGCAGCTCCCGCCAGGAGTCCCTCCACAGTGGCGACCGCTACAAAGGAGTAAGG 642
QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTrp-----SerGlnPro 213
Db 643 GTCCGCTGCTGGAGTCCGCGACCTGCGCGCTCTACACGTTGGGCGGCGGACGTGCC 702
QY 214 GlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTrpProGlu 233
Db 703 CAGGCTGAGCGCAT-----GTGCTGCTGGGAGTCTGTGCTGGCGGTACCCCGAG 753
QY 234 GlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGly 253
Db 754 GGGCACAAGGACGCTGCGAGGGTGAATCTGGGGGACCTCTGACCTGCTGCGGTCTGG 813
QY 254 ArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgArgAsnArgPro 273
Db 814 AGCTGGGTCCTGGTGGGCGTGGTGGAGCTGGGGGAGGGTGTGTCCTGCCCAACCGTCA 873
QY 274 GlyValPheThrAlaValAlaThrTrpGluAlaTrpIleArg----- 287
Db 874 GGGGTCTACACGAGTGTGGCCACATATAGCCCTGGATTTCAGGCTGCGGTCTGCTAA 933
QY 288 -----GluGlnValMetGlySer-----Glu 294
Db 934 TGCTAGCCGCTGAGGCTGACCTGGAGCCAGCTGCTGGGGTCCCTCAGCCTCCTGTTCTAT 993
QY 295 ProGlyProAla---PheProThr 301
Db 994 CCAGGACCTGCTATACCCACCA 1017

Search completed: February 25, 2004, 22:14:22
Job time: 414 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 18:03:06 ; Search time 3777 Seconds

(without alignments)
4096.760 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 357

Sequence: 1 MAQKGVLPGLGAVANSDS.....TKSLVLPWLSPHSLGLMGF 357

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6934743

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10037417/runat_24022004_101108_18083/app_query_fasta_1.519
-DB=GenEmbl -QFMT=fastap -SUFFIX=olip2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417@cgn_1_4958@runat_24022004_101108_18083 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rat.*
- 36: em.htg.rod.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htg.hum.*
- 40: em.htg.mus.*
- 41: em.htg.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	%		
1	357	100.0	1102	6	AX512287	AX512287 Sequence
2	357	100.0	1102	6	AX512289	AX512289 Sequence
3	272	76.2	2457	6	AX360096	AX360096 Sequence
4	272	76.2	2681	6	AX480935	AX480935 Sequence
5	261	73.1	2810	6	BD127529	BD127529 Primer fo
6	261	73.1	2810	9	AK075142	AK075142 Homo sapi
7	178	49.9	537	6	AX342934	AX342934 Sequence
8	132	37.0	670	6	BD125219	BD125219 Primer fo
9	132	37.0	670	6	BD126485	BD126485 Primer fo
10	128	35.9	456	6	AX342936	AX342936 Sequence
11	93	26.1	127769	9	AC009088	AC009088 Homo sapi
12	93	26.1	195476	2	AC135044	AC135044 Homo sapi
13	35	9.8	232119	2	AC106629	AC106629 Rattus no
14	35	9.8	244161	2	AC117170	AC117170 Rattus no
15	29	8.1	188064	9	AC093520	AC093520 Homo sapi
16	23	6.4	64328	2	AC124816	AC124816 Mus muscu
17	23	6.4	185788	10	AC124461	AC124461 Mus muscu
18	23	6.4	188559	2	AC124602	AC124602 Mus muscu
19	23	6.4	219763	10	AC093175	AC093175 Mus muscu
20	19	5.3	71177	2	AC101465	AC101465 Mus muscu
21	17	4.8	6207	9	BC035589	BC035589 Homo sapi
22	17	4.8	7604	9	AB002294	AB002294 Human rRN
23	17	4.8	182230	9	AC135050	AC135050 Homo sapi
24	16	4.5	1161	6	AX675581	AX675581 Sequence
25	16	4.5	1726	6	AX675579	AX675579 Sequence
26	16	4.5	1796	6	AX098193	AX098193 Sequence
27	16	4.5	1800	9	BC001462	BC001462 Homo sapi
28	16	4.5	1834	6	AX335777	AX335777 Sequence
29	16	4.5	1834	6	AX336076	AX336076 Sequence
30	16	4.5	1834	6	AX336340	AX336340 Sequence
31	16	4.5	1834	6	AX409567	AX409567 Sequence
32	16	4.5	1834	6	AX474697	AX474697 Sequence
33	16	4.5	1834	9	HUMPRQS	L41351 Homo sapien
34	16	4.5	1835	6	AX098215	AX098215 Sequence
35	16	4.5	7008	9	HSU33446	U33446 Human prost
36	16	4.5	71177	2	AC101465	AC101465 Mus muscu
37	16	4.5	195476	2	AC135044	AC135044 Homo sapi
38	14	3.9	1931	10	MMX99571	X99571 M.musculus
39	14	3.9	1963	10	BC037085	BC037085 Mus muscu
40	14	3.9	1968	10	BC057921	BC057921 Mus muscu
41	14	3.9	1980	10	BC049867	BC049867 Mus muscu
42	13	3.6	506	6	AR263894	AR263894 Sequence
43	13	3.6	603	10	AB074296	AB074296 Mus muscu
44	13	3.6	711	6	AX763030	AX763030 Sequence
45	13	3.6	792	6	AR337567	AR337567 Sequence

ALIGNMENTS

```
AX512287
LOCUS AX512287 1102 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 43 from Patent WO02053742.
ACCESSION AX512287
VERSION AX512287.1 GI:23392631
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Kekuda,R., Alsobrook,J.P., Tchiernev,V.T., Liu,X., Spytek,K.A.,
Patturajan,M., Grosse,W.M., Lepley,D.M., Burgess,C.E., Vernet,C.A.,
Li,L., Gorman,L., Edinger,S., Sciore,P., Ellerman,K., Malyankar,U.,
Rothenberg,M., Stone,D., Boldog,F., Shenoy,S. and Anderson,D.
Proteins and nucleic acids encoding same
TITLE Patent: WO 02053742-A 43 11-JUL-2002;
JOURNAL Curagen Corporation (US)
FEATURES
source
Location/Qualifiers
1..1102
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1102
Score: 357.00 Matches: 357
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-037-417-46 (1-357) x AX512287 (1-1102)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAAsnSerAspSer 20
DB 19 ATGGCCAGAGGGGTCTCTGGGCTGGGAGCTGGGGCTGGTGGCCAAATTCGACTCA 78
QY 21 TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
DB 79 TACTCACTTTACGGGTGTGGCTGGAGCCGGCTAGGGGGCCCCCGTACTGCGGGCC 138
QY 41 ProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
DB 139 CCTGAGCCCTCGGCCCGCATCTGGGGGCTCAACGCGAGCGGGCACCTGGCCTTG 198
QY 61 GlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSer 80
DB 199 CAAGTAGCCCTGCACCATGAGGTGGCCACATCTGCGGGGGCTCCCTCATCGCCCTCC 258
QY 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
DB 259 TGGGTCTCTCCGCTGCTCACTGTTTCATGACGATGGAGCGTGGAGCCCGGGCGAG 318
QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
DB 319 TGGTGGTACTGTGGGGCTGCATCTCCAGACGGGGCCCCCTGGAGCGGGCGCACCCGC 378
QY 121 AlaValAlaAlaIleValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
DB 379 GCAGTGGCGGCATCGTGGTCCCGCCCAACTACAGCAAGTGGAGCTGGGGCGGCACCTG 438
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160
DB 439 GCCCTGTGCGCTGGCCCTACCCCGCCAGCTGGGCCCGCCGCTGGCTGTCTGCTG 498
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
DB 499 CCGCGGCTCACACCGCTTGTGACAGCGCACCGCTGCTGGGCCACCGGCTGGGGAG 558
QY 181 ValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeu 200
```

```
DB 559 GTCCAGAGGCGAGATCCTCTGCCTCTCCCTCGGTCTACAGGAAGTGAGCTAAGGCTG 618
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
DB 619 CTGGGCGAGGCCACCTGTCAATGCTCTACAGCAGCCCGTCCCTTCACTCACTCTC 678
QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgAspThrCysGln 240
DB 679 CAGATATTGCCAGGATGCTGTGTGTGGCTACCCAGAGGCCCGCAGGACACCTGCCAG 738
QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGlnAlaGlyIle 260
DB 739 GGTGACTCTGGGGGGCCCCCTGGTCTGTGAGAAAGGCGCGCTGGTTCCAGCAGGAATC 798
QY 261 ThrSerPheGlyPheGlyCysGlyArgAsnArgProGlyValPheThrAlaValAla 280
DB 799 ACCAGCTTTGGTTGGCTGTGGACGAGAAACCGCCCTGGAGTTTTCACCTGCTGGCT 858
QY 281 ThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyProAlaPhePro 300
DB 859 ACCTATGAGCATGGATACGGGAGCAGGTATGGTTTCAGAGCCTGGGCTGCTTTCCC 918
QY 301 ThrGlnProGlnLysThrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320
DB 919 ACCAGCCCCCAGAGACCCAGTCAGATTGTTTACATCAACCGCATTCCTGGATTCTGCC 978
QY 321 ArgIleLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrIysSer 340
DB 979 AGAATCTCTTTGAGGCCCTTGTCCCATATATCATAGTAGAGTCTCACTGGGACCAAGC 1038
QY 341 LeuValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
DB 1039 CTTGTCTCTCTCTGGCTCTCTCCACACTCTCTCTGGGCTCTGGGGGTTT 1089
RESULT 2
AX512289
LOCUS AX512289 1102 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 45 from Patent WO02053742.
ACCESSION AX512289
VERSION AX512289.1 GI:23392632
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Kekuda,R., Alsobrook,J.P., Tchiernev,V.T., Liu,X., Spytek,K.A.,
Patturajan,M., Grosse,W.M., Lepley,D.M., Burgess,C.E., Vernet,C.A.,
Li,L., Gorman,L., Edinger,S., Sciore,P., Ellerman,K., Malyankar,U.,
Rothenberg,M., Stone,D., Boldog,F., Shenoy,S. and Anderson,D.
Proteins and nucleic acids encoding same
TITLE Patent: WO 02053742-A 45 11-JUL-2002;
JOURNAL Curagen Corporation (US)
FEATURES
source
Location/Qualifiers
1..1102
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1102
Score: 357.00 Matches: 357
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-037-417-46 (1-357) x AX512289 (1-1102)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAAsnSerAspSer 20
DB 19 ATGGCCAGAGGGGTCTCTGGGCTGGGAGCTGGGGCTGGTGGCCAAATTCGACTCA 78
QY 21 TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
DB 79 TACTCACTTTACGGGTGTGGCTGGAGCCGGCTAGGGGGCCCCCGTACTGCGGGCC 138
QY 41 ProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
DB 139 CCTGAGCCCTCGGCCCGCATCTGGGGGCTCAACGCGAGCGGGCACCTGGCCTTG 198
QY 61 GlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSer 80
DB 199 CAAGTAGCCCTGCACCATGAGGTGGCCACATCTGCGGGGGCTCCCTCATCGCCCTCC 258
QY 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
DB 259 TGGGTCTCTCCGCTGCTCACTGTTTCATGACGATGGAGCGTGGAGCCCGGGCGAG 318
QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
DB 319 TGGTGGTACTGTGGGGCTGCATCTCCAGACGGGGCCCCCTGGAGCGGGCGCACCCGC 378
QY 121 AlaValAlaAlaIleValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
DB 379 GCAGTGGCGGCATCGTGGTCCCGCCCAACTACAGCAAGTGGAGCTGGGGCGGCACCTG 438
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160
DB 439 GCCCTGTGCGCTGGCCCTACCCCGCCAGCTGGGCCCGCCGCTGGCTGTCTGCTG 498
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
DB 499 CCGCGGCTCACACCGCTTGTGACAGCGCACCGCTGCTGGGCCACCGGCTGGGGAG 558
QY 181 ValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeu 200
```

QY 21 TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
 Db 79 TACTACATTTACGGTTGGTGGCGTCGTCGACCGCTAGGGGCCCCCGTACTGGGGCGC 138
 QY 41 ProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
 Db 139 CTGAGCCCTCGGCCCGCATCGTGGGGGGCTCAACGCGCAGCGCGGCACCTGGCCCTGG 198
 QY 61 GlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIleAlaProSer 80
 Db 199 CAAGTGAGCTGCACCATGAGGTGGCCCATCTGCGGGGGCTCCCTCATCGGCCCTCC 258
 QY 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
 Db 259 TGGGTCTCTCCGCTGCTCACTGTTCATGACGANTGGGAGCTGTGGAGCCCGGCGGAG 318
 QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
 Db 319 TGGTGGTACTGCTGGCGGTGCACCTCCAGGACGGGCCCCGTGGACGGCGGCACCCGC 378
 QY 121 AlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
 Db 379 GCAGTGGCCGCCATCGGTGGCGGCCCACTACAGCCAGTGGAGCTGGGGCGCGACCTG 438
 QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160
 Db 439 GCCCTGCTGGCCCTGGCTCACCGCAGCTGGGCCCGCCGCTGGCTGTGTGCTGCTG 498
 QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
 Db 499 CCGCGCCCTCACACCGCTCTGTGACGGCACCCGCTGCTGGGCCACCGGCTGGGGAGAC 558
 QY 181 ValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnValGluLeuArgLeu 200
 Db 559 GTCCAGAGGAGATCTCTGCTCTCCCTGGGTGCTACAGGAGTGGAGCTAGAGCTG 618
 QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
 Db 619 CTGGGCGAGGCCACTGTCAATGTCTCTACAGCAGCCCGCTCCCTTCACTCACTCTC 678
 QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240
 Db 679 CAGATATTGCCAGGAGTCTGTGTGCTGGCTACCCAGAGGGCCGACAGGCACCTGCCAG 738
 QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyArgTrpPheGlnAlaGlyIle 260
 Db 739 GGTCACTTGGGGGGCCCTGCTGTGAGGAGCGGCCCTGCTTCCAGCAGGAATC 798
 QY 261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
 Db 799 ACCAGCTTGGTTGGCTGTGGAGCGGAGAAACCGCCCTGGAGTTTTCACCTGCTGTG 858
 QY 281 ThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyProAlaPhePro 300
 Db 859 ACCTATGAGGATGATACGGGAGCAGGTGATGGGTTCAGAGCCTGGGCCCTGCCCTTCCC 918
 QY 301 ThrGlnProGlnLysThrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320
 Db 919 ACCAGCCCCAGAACCCAGTCAGATTGTTTACATCAACCGCATTCCTGGATTCTGCC 978
 QY 321 ArgIleLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrLysSer 340
 Db 979 AGAATCTTTTGGGCCCTTGTCCCATATATCAGTAGGAGTCTCACTGGGACCAAAAGC 1038
 QY 341 LeuValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
 Db 1039 CTGTGCTCTCCCTGGCTCTCTCCACACTCTCTCTGGGCCCTCTGGGGGTTC 1089

RESULT 3

AX360096 2457 bp DNA linear PAT 13-FEB-2002
 LOCUS
 DEFINITION Sequence 52 from Patent WO200860.

ACCESSION AX360096 GI:18675722
 VERSION AX360096.1
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Plowman, G., Whyte, D., Sudarsanam, S., Manning, G., Caenepeel, S. and Charydzak, G.
 TITLE Novel proteases
 JOURNAL Patent: WO 0200860-A 52 03-JAN-2002;
 FEATURES
 source Location/Qualifiers
 1..2457
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores: Length: 2457
 Pred. No.: 241e-265
 Score: 272.00
 Matches: 272
 Percent Similarity: 100.00%
 Conservative: 0
 Best Local Similarity: 100.00%
 Mismatches: 0
 Query Match: 76.19%
 Indels: 0
 Gaps: 6
 DB: 6
 US-10-037-417-46 (1-357) x AX360096 (1-2457)
 QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
 Db 112 TGGGGCGCCCTGAGCCCTCGGCCCGCATCTGTTGGGGGCTCAAAACGGCGACCGGGGACC 171
 QY 58 TrpProTrpGlnValSerLeuHisGlyGlyHisIleCysGlyGlySerLeuLeu 77
 Db 172 TGGCCTTGGCAAGTGGAGCTGCACCATGGAGTGGCCACATCTGGGGGGCTCCCTCATC 231
 QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
 Db 232 GCCCCTCTCTGGTCTCTCCGCTGCTCACTGTTCATGACGAATGGGACGCTGGAGCCC 291
 QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
 Db 292 GCGGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 351
 QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGly 137
 Db 352 CACACCGCGCAGTGGCGCCCATCTGCTGGCGCGTGCATCCCGAGACGGGCCCTTGGAGCGCGC 411
 QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
 Db 412 GCCACCTGGCCCTGCTGGCCTGCGCTCAACCGCTTGTGACCGCACCGCTGCTGGCGCCG 471
 QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
 Db 472 GTCTGCTTGGCCCCCGCTCACACCGCTTGTGACCGCACCGCTGCTGGCGCACCGGC 531
 QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnValGlu 197
 Db 532 TGGGGAGAGCTCCAGGAGGAGATCTCTGCTCTCCCTGGGTGTCTACAGGAAGTGGAG 591
 QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
 Db 592 CTAAGGCTGCTGGGCGAGGCCACCTGTCAATGTCTCTACAGCCAGCCCGGCTCCCTTCAAC 651
 QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyValArgArgAsp 237
 Db 652 CTCACCTTCCAGATATTGCCAGGAGATGTGTGTGCTGTGCTACCCAGAGGGCGGAGGAGAC 711
 QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGlyGlyArgTrpPheGln 257
 Db 712 ACCTGCCAGGCTGACTCTGGGGGGCCCCCTGTGTGTGAGGAAGGCGCGCTGGTTCCAG 771

QY 258 AlaGlyThrSerPheGlyPheGlyCysGlyArgGlnArgProGlyValPheThr 277
 Db 772 GCAGGAATCACCAGCTTTGGCTTTGGCTGTGGAGCGGAGAACCGCCCTGGAGTTTCACT 831

QY 278 AlaValAlaThrTyrGluAlaTyrPheGlyValMetGlySerGluProGlyPro 297
 Db 832 GCTGTGGCTACTATGAGGAGATGATATACGGAGAGAGGTGATGGTTTCAGAGCTGGGCT 891

QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp 309
 Db 892 GCCTTTCCACCCAGCCAGAGACCCAGACCCAGTCCAGAT 927

RESULT 4
 AX480935 2681 bp DNA linear PAT 12-AUG-2002
 LOCUS Sequence 31 from Patent WO0246383.
 DEFINITION AX480935
 ACCESSION AX480935
 VERSION AX480935.1 GI:22217574
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Yue, H.; Aizumi, Y.; Kallik, D.A.; Baughn, M.R.; Griffin, J.A.; Swarnakar, A.; Lai, P.G.; Walla, N.K.; Hafalia, A.J.; Gandhi, A.R.; Au-Young, J.; Elliott, V.S.; Ramkumar, J.; Thangavelu, K.; Lu, Y.; Warren, B.A.; Lu, D.A.; Lee, E.A.; Tribouley, C.M.; Arvizu, C.; Deleane, A.M.; Yao, M.G.; Khan, F.A. and Sanjanwala, M.M.
 TITLE Protein modification and maintenance molecules
 JOURNAL Patent: WO 0246383-A 31 13-JUN-2002;
 Incyte Genomics, Inc. (US)
 FEATURES
 Location/Qualifiers
 1..2681
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 2751509CB1"

ORIGIN
 Alignment Scores: 2,57e-265 Length: 2681
 Pred. No.: 272.00 Matches: 272
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 76.19% Indels: 0
 Query Match: 6 Gaps: 0
 DB:

US-10-037-417-46 (1-357) x AX480935 (1-2681)

QY 38 CysGlyArgProGluProSerAlaArgTyrGlySerAsnAlaGlnProGlyThr 57
 Db 112 TGGCGGCGCCCTCAGCCCTCGGCGCATCGTGGGGGCTCAACCGCGAGCGGGGACC 171

QY 58 TrpProThrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeuIle 77
 Db 172 TGGCGCTTGGCAAGAGCGCTGCACCATGGAGTGGCCACATCTGCGGGGCTCCCTCATC 231

QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
 Db 232 GCCCCTCTCGGCTCTCCGCTGCTCACTGTTTCATGCAATGAGGAGCGCTGGAGCC 291

QY 98 AlaAlaGluTrpSerValLeuGlyValHisSerGlnAspGlyProLeuAspGlyVal 117
 Db 292 GCGGCGGAGTGGTGGTACTCTGCGGGGTGACATCCAGGAGCGGCGGCTGGAGCGGCG 351

QY 118 HisThrArgAlaValAlaAlaIleValProAlaSerTyrSerGlnValGluLeuGly 137
 Db 352 CACACCCCGCAGTGGCGCCCATCTGTGTCGCGCCACTACAGCCAGTGGAGCTGGG 411

QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157

Db 412 GCGCACTGGCCCTGCTGGCCCTGGCTCACCAGCCAGCGCTGGGCGCCGCTGGCT 471

QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
 Db 472 GTCTGCTGCCCGCGCTCACCAGCTTCGTCACGCGCAGCCGCTGCTGGGCGCAGCGC 531

QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
 Db 532 TGGGAGACGCTCCAGGAGGAGATCCTCTGCTCTCCCTTGGGTGTACAGGAGTGGAG 591

QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
 Db 592 CTAAGGCTGTGGCGGAGGAGCCACTGTCAATGCTCTACAGCCAGCCCGCTCCCTTCAAC 651

QY 218 LeuThrLeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
 Db 652 CTCACTCTCCAGATATTGCCAGGAGTCTGTGTGCTTACCCAGAGGCGCGAGGAGAC 711

QY 238 ThrCysGlnGlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGln 257
 Db 712 ACCTGCCAGGAGTACTCTGGGGGCGCTGTGTGTGAGGAGGCGGCGCTGTGTTCAG 771

QY 258 AlaGlyThrSerPheGlyPheGlyCysGlyArgGlnArgProGlyValPheThr 277
 Db 772 GCAGGAATCACCAGCTTTGGCTTTGGCTGTGGACGGAGAACCCGCTGGAGTTTCACT 831

QY 278 AlaValAlaThrTyrGluAlaTyrPheGlyValMetGlySerGluProGlyPro 297
 Db 832 GCTGTGGCTACTATGAGGAGTATGATATGAGGAGGAGGAGTGTGGTTTCAGAGCTGGGCT 891

QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp 309
 Db 892 GCCTTTCCACCCAGCCAGAGACCCAGTCCAGAT 927

RESULT 5
 BD127529 2810 bp DNA linear PAT 18-SEP-2002
 LOCUS Primer for synthesizing full-length cDNA and use thereof.
 DEFINITION BD127529
 ACCESSION BD127529
 VERSION BD127529.1 GI:23222474
 KEYWORDS JP 2002017375-A/2960.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2810)
 AUTHORS Ota, T.; Nishikawa, T.; Isogai, T.; Hayashi, K.; Ishii, S.; Kawai, Y.; Wakamatsu, A.; Sugiyama, T.; Nagai, K.; Kojima, S.; Otsuki, T. and Koga, H.
 TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL Patent: JP 2002017375-A 2960 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 COMMENT OS Homo sapiens (human)
 PN JP 2002017375-A/2960
 PD 22-JAN-2002
 PF 07-JUN-2000 JP 2000253172
 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO ISHII,
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI SHINICHI KOJIMA,
 PI TETSUJI OTSUKI, HISASHI KOGA
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,
 C12N15/00, C12N15/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH key
 Location/Qualifiers
 1..2810
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

FT CDS Location/Qualifiers
 1..2810
 (300)..(2597).

AK075142.1 GI:22761040
 VERSION
 oligo capping: fis (full insert sequence).
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 REFERENCE
 AUTHORS
 Isoqai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T., Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S., Kawai-Hio, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Kojima, S., Nagatani, K., Masuho, Y., Ono, T., Okano, K., Yoshikawa, Y., Aotsuka, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and Ninomiya, K.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 2810)
 Isogai, T. and Otsuki, T.
 Direct Submission
 Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; cDNA 5' & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).
 Location/Qualifiers
 1..2810
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PLACE1005003"
 /tissue_type="placenta"
 /clone_lib="PLACE1"
 /note="cloning vector: pME18SFL3"
 300..2600
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAC11431.1"
 /db_xref="GI:22761041"
 /translation="MTNGTLEPAEWSVLLGVHSDQPLDGAFTRAVAIVVPANYSQVILGADALLRLASPASIGPVPVCLPRASHRFVHGTAQWATGMDGVQDADPLPLPVLQVEVLRLIGATCCQLYSQGPENLTQILQMLCAGYPEGRRDTCCGDSGGLVCIEGGRWFAGITTSFGCGRRNRPVFTAVAYEAWIRQVMSGEPAPFTQPKQTSQPFREBENTIALPECKAPRPPGAPWAEQVMPGSRPCGALVSWLRLAPASCFIDPNSDPPDRDLAWRVLLPSRPABRVARLVHENSADNSDALIALLQRLTPMVLASRPYCLPFPHPHYFLPGSRCLARWGEPALPGALLAEILLGWMCHCLYGRQGAAPVLPDGPALCPAYOEKEVSGSCWNSRWSLLQCEGTWFLAGIRWDFPSCGLRPAFPLOTHGPMI SHVTGRAYLDQAWMDPGDGETETQCPHTEHGAACGLRLAEPVFWLWPLAEVHVAGDRCVTGILLACGMWLAATHCVLRPGSTVTPYIEVLGRAGASLLPQHQVSRVLSIRLPHQLGRPLPALLLELSRVEPSPALPCLHPAGIPPGASGCVLWMTGPDQVPAVAASLIITQRCDCLYQIILPPGTLVLYAEGQNRCEMTSAPPELLCQMTGEGSWLVGVMAVQGSRELFPAIIGPEAWISQTVCEANLPPSGSPHWPFTGGSNLPPELAKSGSPHAVYFLLLLTLLIQS"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,78e-254 Length: 2810
 Score: 261.00 Matches: 261
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 73.11% Indels: 0
 DB: 9 Gaps: 0
 US-10-037-417-46 (1-357) x AK075142 (1-2810)
 Qy 49 GlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisGlyGly 68
 Db 177 GGGGGCTCAAAACCGGACCGCGGACCTGGCTTGGCAGTGAGCTGCACCATGGAGT 236
 Qy 69 GlyHisIleCysGlyGlySerLeuAlaProSerTrpValLeuSerAlaHisCys 88
 Db 237 GGGCACATCGCGGGGCTCCCTCATCGCCCTCTCTGGTCTCTCCGCTGCTCACTGT 296
 Qy 89 PheMetThrAsnGlyThrLeuGluProAlaAlaGlnTrpSerValLeuGlyValHis 108
 Db 297 TTTATGACGAATGGAGCGCTGGAGCCCGCGCGAGTGGTGGTACTGCTGGCGGTGCAC 356
 Qy 109 SerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaLleValPro 128
 Db 357 TCCAGGACGGGCCCTCGGACGGCGCACACCGCGCAGTGGCGCGCATCGTGTGGCGG 416
 Qy 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuAlaSerPro 148
 Db 417 GCCAACTACAGCAAGTGGAGCTGGGCGCGACCTGGCCCTCTGGCCCTTGACCC 476
 Qy 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168
 Db 477 GCGAGCTGGCGCGCGTGTGGCTGTCTGCTGCGCCCGCGCTCACACCGCTTCTGTG 536
 Qy 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAlaAspProLeuPro 188
 Db 537 CACGGCACCGCTGTGGGCCACCGGCTGGGAGACGCTCCAGGAGGAGCATCTCTCTGCCT 596
 Qy 189 LeuProTrpValLeuGlnValGluLeuArgLeuLeuGlyGluAlaThrCysGlnCys 208
 Db 597 CTCCTCTGGGTCACAGAAAGTGAGCTAAGCTGTCTGGCGAGGCCACCTGTCAATGT 656
 Qy 209 LeuTyrSerGlnProGlyProPheAsnLeuThrLeuGlnIleLeuProGlyMetLeuCys 228
 Db 657 CTCTACAGCCAGCCGGTCCCTTCAACCTCATCTCTCCAGATATTGCCAGGGAGATCTGTGT 716
 Qy 229 AlaGlyTyrProGlyGlyArgArgAspThrCysGlnGlyAspSerGlyGlyProLeuVal 248
 Db 717 GCTGCTACCCAGAGGGCGCGAGGACACTGGCCAGGCGGACTCTGGGGGGCCCCCTGGTC 776
 Qy 249 CysGluGluGlyGlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGly 268
 Db 777 TGTGAGGAAGCGCGCGCTGGTTCCAGGACAGAAATCACACAGCTTTGGCTTGTGGA 836
 Qy 269 ArgArgAsnArgProGlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGlu 288
 Db 837 CGGAGAAACCGCCCTGGAGTTTTTCACTGTGGTACTCTATGAGGCATGATACGGAG 896
 Qy 289 GlnValMetGlySerGluProGlyProAlaPheProThrGlnProGlnLysThrGlnSer 308
 Db 897 CAGGTGATGGGTTCAGAGCTGGGCTGGCTTCCACCCAGCCCGACAGACCCAGTCA 956
 Qy 309 Asp 309
 Db 957 GAT 959
 RESULT 6
 AK075142
 LOCUS
 DEFINITION
 Homo sapiens cDNA FLJ90661 fis, clone PLACE1005003, weakly similar to PROSTASIN PRECURSOR (EC 3.4.21.-).
 2810 bp mRNA linear PRI 03-SEP-2002

COMMENT	OS	Homo sapiens (human)
PN	JP 2002017375-A/1916	
PD	22-JAN-2002	
PF	07-JUL-2000 JP 2000253172	
PI	TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO	
PI	ISHII,	
PI	YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI	
SHINICHI KOJIMA,		
PI	TETSUJI OTSUKI, HISASHI KOGA	
PC	C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC	
PC	Primer for synthesizing full-length cDNA and use thereof FH	Key
FT	source	1. .670
FT	Location/Qualifiers	
FEATURES	source	1. .670
ORIGIN	Location/Qualifiers	/organism="Homo sapiens (human)";
Alignment Scores:		
Pred. No.:	8.13e-124	Length: 670
Score:	132.00	Matches: 132
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	36.97%	Indels: 0
DB:	6	Gaps: 0
US-10-037-417-46 (1-357) x BD125219 (1-670)		
QY	49 GlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHisGlyGly	68
DB	177 GGGGGCTCAACGCGCAGCGCGGACCTGGCCCTTGGCAAGTGGAGCTGCACCATGGAGGT	236
QY	69 GlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCys	88
DB	237 GGGCCACATCTGGGGGGCTCCCTCATCGCCCTCTCTGGGTCTCTCGCTGCTCACTGT	296
QY	89 PheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHis	108
DB	297 TTCATGAGCAATGGAGCGCTGGAGCGCGCGGCACCTGGTACGTCTGCTGGCGTGCAC	356
QY	109 SerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValPro	128
DB	357 TCCAGAGCGGGCCCTGGAGCGCGCGGCACACCGCGCAGTGGCGCATCTGCTGGCGG	416
QY	129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerPro	148
DB	417 GCCAACTACAGCAAGTGGAGCTGGCGCGCGACCTGGCCCTGCTGGCGCTGCACCC	476
QY	149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal	168
DB	477 GCACGCTGGGCCCCGGTGGCTGTCTGCTGCGCCCGCGCGCTGCACCGCTTGTG	536
QY	169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp	180
DB	537 CACGGCACCGCTGCTGGGCCACCGCTGGGAGAC	572
RESULT 9		
BD126485	670 bp	DNA linear PAT 18-SEP-2002
LOCUS	Primer for synthesizing full-length cDNA and use thereof.	
DEFINITION	BD126485	
ACCESSION	BD126485.1	GI:23221430
VERSION	JP 2002017375-A/1916.	
KEYWORDS	Homo sapiens (human)	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 670)	
AUTHORS	Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y., Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and Koga, H.	
TITLE	Primer for synthesizing full-length cDNA and use thereof	
JOURNAL	Patent: JP 2002017375-A 1916 22-JAN-2002; HELIX RESEARCH INSTITUTE	

REFERENCE 1
 AUTHORS Xiao,Y. and Morozov,V.
 TITLE Regulation of human prostatic-like serine protease
 JOURNAL Patent: WO 0198467-A 3 27-DEC-2001;
 BAYER Aktiengesellschaft (DE)
 FEATURES Location/Qualifiers
 source 1..456
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Alignment Scores: 6.98e-120 Length: 456
 Pred. No.: 128.00 Matches: 128
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 35.85% Gaps: 0
 DB: 6

US-10-037-417-46 (1-357) x AX342936 (1-456)

QY 121 AlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
 DB 384 GCAGTGGCGGCATCGTGTGGCGCCAACTACAGCAAGTGGAGCTGGCGCGGACCTG 325

QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyrProValCysLeu 160
 DB 324 GCCTGTCTGGCGCTGCTCCACCGCCGACCTGGCGCGCGCTGTCTGCTG 265

QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTyrAlaThrGlyTyrGlyAsp 180
 DB 264 CCGCGCGCTCACACCGCTTCTGTGACGCGCACCGCTGTGGCCACCGCTGGGAGAC 205

QY 181 ValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGluValGluLeuArgLeu 200
 DB 204 GTCCAGGAGCGAGATCTCTGCTCTCTCCCTGCTGTACAGGAGTGGAGCTAAGGCTG 145

QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
 DB 144 CTGGCGGAGCGCACCTGTCAATGTCTACAGCACCGCGCTGTGTCAACCTCACTCTC 85

QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyValArgAspThrCysGln 240
 DB 84 CAGATATTGCCAGGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 25

QY 241 GlyAspSerGlyGlyProLeuVal 248
 DB 24 GGTGACTCTGGGGGGCCCTGGTC 1

RESULT 11
 AC009088/c 127769 bp DNA linear PRI 29-MAR-2003
 LOCUS Homo sapiens chromosome 16 clone RP11-388M20, complete sequence.
 DEFINITION AC009088
 ACCESSION AC009088.9 GI:29366934
 VERSION HTG.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 127769)
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 127769)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 127769)
 AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission
 JOURNAL Submitted (03-APR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 4 (bases 1 to 127769)
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Submitted (18-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 5 (bases 1 to 127769)
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Mar 29, 2003 this sequence version replaced gi:29029216.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
 www.sbgc.stanford.edu
 Quality: Phrap Quality >=40 99.7% of Sequence;
 Estimated Total Number of Errors is 0.2.
 NOTE: This insert is not the entire sequence of the clone (entire sequence is 233.4kb). It is clipped at the overlap with AC135050.
 The number of bases overlapped is 2575.

FEATURES
 Location/Qualifiers
 1..127769
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-388M20"

ORIGIN
 Alignment Scores: 1.31e-82 Length: 127769
 Pred. No.: 93.00 Matches: 93
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 26.05% Gaps: 0
 DB: 9

US-10-037-417-46 (1-357) x AC009088 (1-127769)

QY 92 AsnGlyThrLeuGluProAlaAlaGluTyrSerValLeuLeuGlyValHisSerGlnAsp 111
 DB 31589 AATGGAGCGCTGGAGCCCGCGCGAGTGGTGGTACTCTGGCGGTGCACCTCCAGGAC 31530

QY 112 GlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaIleValValProAlaAsnTyr 131
 DB 31529 GGGCCCCCTGGAGCGCGCGCACACCGCGCGAGTGGCGCCATCTGTGTGTGGCGCACTAC 31470

QY 132 SerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeu 151
 DB 31469 AGCCAACTGGAGCTGGCGCGCGAGCTGGCGCTGTGGCTGTGGCTGTGGCTGTGGCTGT 31410

QY 152 GlyProAlaValTyrProValCysLeuProArgAlaSerHisArgPheValHisGlyThr 171
 DB 31409 GGGCCCCCGCGTGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 31350

QY 172 AlaCysTyrAlaThrGlyTyrGlyAspValGlnGluAla 184
 DB 31349 GCCTGTGGCGCCACCGCGCTGGGGAGAGCTGCCAGGAGCA 31311

RESULT 12
 AC135044 195476 bp DNA linear HTG 05-OCT-2002
 LOCUS Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
 DEFINITION AC135044
 ACCESSION AC135044
 VERSION AC135044.1 GI:23505535
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE
ORGANISM

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 195476)
DOE Joint Genome Institute.

TITLE
JOURNAL

Sequencing of Human Chromosome 16
Unpublished

REFERENCE
AUTHORS

2 (bases 1 to 195476)
DOE Joint Genome Institute.
Direct Submission
Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

TITLE
JOURNAL

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

COMMENT

Project Information
Center Project Name: 809609
Center clone name: C17B-EL_2551B20

Summary Statistics

Consensus quality: 171229 bases at least Q40
Consensus quality: 182638 bases at least Q30
Consensus quality: 188095 bases at least Q20
Estimated insert size: 170000; agarose-fp estimation
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1202: contig of 1202 bp in length
1203 1302: gap of unknown length
1303 2468: contig of 1166 bp in length
2469 2568: gap of unknown length
2569 4077: contig of 1509 bp in length
4078 4177: gap of unknown length
4178 5772: contig of 1595 bp in length
5773 5872: gap of unknown length
5873 7722: contig of 1850 bp in length
7723 7822: gap of unknown length
7823 8958: contig of 1136 bp in length
8959 9058: gap of unknown length
9059 11051: contig of 1993 bp in length
11052 11151: gap of unknown length
11152 13641: contig of 2490 bp in length
13642 13741: gap of unknown length
13742 17738: contig of 3997 bp in length
17739 17838: gap of unknown length
17839 22030: contig of 4192 bp in length
22031 22130: gap of unknown length
22131 27002: contig of 4872 bp in length
27003 32819: contig of 5717 bp in length
32820 32919: gap of unknown length
32920 39003: contig of 6084 bp in length
39004 39103: gap of unknown length
39104 48370: contig of 9267 bp in length
48371 48470: gap of unknown length
48471 59821: contig of 11251 bp in length
59822 59821: gap of unknown length
59823 72404: contig of 12483 bp in length
72405 72405: gap of unknown length
72406 85349: contig of 12945 bp in length
85350 85350: gap of unknown length
85450 101872: contig of 16423 bp in length

* 101873 101972: gap of unknown length
* 117076: contig of 15104 bp in length
* 117077 117176: gap of unknown length
* 117177 138973: contig of 21797 bp in length
* 138974 139073: gap of unknown length
* 139074 150336: contig of 11263 bp in length
* 150337 150436: gap of unknown length
* 150437 195476: contig of 45040 bp in length.

FEATURES
source

1. .195476
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="CTD-2551B20"
/clone_lib="Caltech human BAC library D"

ORIGIN

Alignment Scores:
Pred. No.: 1.79e-82 Length: 195476
Score: 93.00 Matches: 93
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.05% Indels: 0
DB: 2 Gaps: 0

US-10-037-417-46 (1-357) x AC135044 (1-195476)

Qy 92 AsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAsp 111
Db 27407 AATGGACGCTGGAGCCGCGCGGAGTGGTCTGCTACTGCGGCGTGCACTCCAGGAC 27466
Qy 112 GlyProLeuAspGlyAlaHisThrArgAlaValAlaAlaLeuValProAlaAsnTyr 131
Db 27467 GGGCCCTCGGACGGCGGCACACCCGCGCAGTGGCGCCATCGTGTGTCGCGGCACTAC 27526
Qy 132 SerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeu 151
Db 27527 AGCCACGTGGAGCTGGCGCGGACCTGGCCCTGCTGGCCCTCACCCGCCGCGCTG 27586
Qy 152 GlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheValHisGlyThr 171
Db 27587 GGGCCCGCCGCTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 27646
Qy 172 AlaCysTrpAlaThrGlyTrpGlyAspValGlnGluAla 184
Db 27647 GCCTGCTGGCGCACCGGCTGGCGGAGACGTCAGGAGGCA 27685

RESULT 13
AC106629/c

LOCUS 232119 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-21802, WORKING DRAFT SEQUENCE, 2
unordered pieces.

AC106629

AC106629.4 GI:30579402
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FUILLTOP.

KEYWORDS
SOURCE

Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE
AUTHORS

1 (bases 1 to 232119)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Y., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Anayalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gregeorgis, E., Geer, K., Gail, R., Grady, M., Guerra, T., Garza, M.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuhewa, L., Louisseg, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Pal, S.,
 Nwakolamleh, O., Okwuonu, G., Olarunpungoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puzoz, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rokey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Shen, H.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smales, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, D., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.
 Direct Submission
 2 (bases 1 to 244161)
 Unpublished
 Worley, K.C.
 Direct Submission
 Submitted (08-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 244161)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 15, 2002 this sequence version replaced gi:23267402.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GVR
 Center clone name: CH230-20P9
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 223430 bases at least Q40
 Consensus quality: 226412 bases at least Q30
 Consensus quality: 228516 bases at least Q20
 Estimated insert size: 226458; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 25619: contig of 25619 bp in length
 * 25620 25719: gap of unknown length
 * 137872 137872: contig of 112153 bp in length
 * 137873 137972: gap of unknown length
 * 137973 240780: contig of 102818 bp in length
 * 240791 240890: gap of unknown length
 * 240891 241970: contig of 1080 bp in length
 * 241971 242070: gap of unknown length
 * 242071 244161: contig of 2091 bp in length.
 ----- Location/Qualifiers
 1. 244161
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-20P9"
 19835..20652
 /note="clone_boundary
 clone_end:Sp6
 site:
 end sequence:BH342986"
 21143..23721
 /note="wgs contig"
 23772..25619
 /note="wgs contig"
 103292..103777
 /note="clone_boundary
 clone_end:T7
 site:
 end sequence:BH342985"
 129838..131159
 /note="wgs end_extension
 clone_end:T7"
 136097..137872
 /note="wgs end_extension
 clone_end:T7"

 ORIGIN
 Alignment Scores:
 Pred. No.: 1,34e-23 Length: 244161
 Score: 35.00 Matches: 35
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.80% Indels: 0
 DB: 2 Gaps: 0
 US-10-037-417-46 (1-357) x ACL11710 (1-244161)

Qy

55

ProGlyThrTrpProTrpGlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGly 74

Db

129868

CCTGCACCTTGGCGGTGGCAGTGAGCTGCATCACGTTGGAGCCACATCTGCGGGGC 129927

Qy

75

SerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCysPhe 89

Db

129928

TCCCTCATCGCCCTTCTCTGGGTCTCTCCGCTGCTCACTGTTTC 129972

AC093520

168064 bp

DNA

linear

PRI 19-MAR-2003

LOCUS

Homo sapiens chromosome 16 clone RP11-120K18, complete sequence.

DEFINITION

AC093520

ACCESSION

AC093520

VERSION

AC093520.4

GI:29124069

HTG.

KEYWORDS

HTG.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 168064)

AUTHORS

DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.

TITLE

Direct Submission

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 168064)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (13-JUN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

3 (bases 1 to 168064)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (13-JUN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

4 (bases 1 to 168064)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (07-JUL-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

5 (bases 1 to 168064)

AUTHORS

DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.

TITLE

Direct Submission

JOURNAL

Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT

On Mar 19, 2003 this sequence version replaced gi:21702833.
Draft Sequence Produced by DOE Joint Genome Institute
www.tgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
www.sngc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.

Qy

312

HisGlnThrAlaPheLeuAepSerAlaArgTlleLeuLeuArgProLeuSerHisIleSer 331

Db

141592

CATCAACGGCATTTCTGTGATTCTGCCAGATCTTTTGGAGCCCTTGCCCATATATCA 141533

Qy

332

ValGlyValSerThrGlyThrLysSer 340

Db

141532

GTAGGAGTCTCAACTGGGACCAAAAGC 141506

Search completed: February 25, 2004, 23:28:10

Job time : 3995 secs

FEATURES

source

1. 168064

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="16"

/clone="RP11-120K18"

ORIGIN

Alignment Scores:

Pred. No.:

1.23e-17

Length: 168064

Score:

29.00

Matches: 29

Percent Similarity:

100.00%

Conservative: 0

Best Local Similarity:

100.00%

Mismatches: 0

Query Match:

8.12%

Indels: 0

DB:

9

Gaps: 0

US-10-037-417-46 (1-357) x AC093520 (1-168064)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 18:01:15 ; Search time 419 Seconds
(without alignments)

3619.586 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 357

Sequence: 1 MAQGVLFQGLGAVANSDS.....TKSLVLPWLSPLHLLGLWGF 357

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6745146

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10037417/runat_24022004_101108_18072/app_query.fasta_1.519
-DB=N Geneseq 29Jan04 -QMT=fastap -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417@cgn_1_885@runat_24022004_101108_18072 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2003as:*
- 8: geneseqn2003bs:*
- 9: geneseqn2003cs:*
- 10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357	100.0	1102	6	Abq93901 Human pro
2	357	100.0	1102	6	Abq93902 Human pro
3	272	76.2	2457	6	Abk31795 DNA encod
4	272	76.2	2681	6	Abq75956 Human PMW
5	261	73.1	2810	4	Aak94500 Human ful
6	178	49.9	537	6	Abk13565 Human pro
7	132	37.0	597	5	Aas69040 DNA encod
8	132	37.0	670	4	Aak92190 Human cDN

9	132	37.0	670	4	AAK93456	Human cDN
10	128	35.9	456	6	ABK13566	Human pro
11	52	14.6	2298	5	AAH71079	DNA encod
12	45	12.6	768	4	AAH78206	Nucleotid
13	17	4.8	8999	5	ABA21236	Human ner
14	16	4.5	1161	6	ABN85393	Human NOV
15	16	4.5	1726	6	ABN85392	Human NOV
16	16	4.5	1733	7	ABT31936	Human bre
17	16	4.5	1796	5	AAF98698	Human ova
18	16	4.5	1834	6	ABS76501	cDNA enco
19	16	4.5	1834	6	ABL67949	Ovary can
20	16	4.5	1834	6	ABL68512	Kidney ca
21	16	4.5	1834	6	ABL68248	Kidney ca
22	16	4.5	1834	6	ABK12241	cDNA enco
23	16	4.5	1834	6	ABN95716	Gene #221
24	16	4.5	1835	5	AAF98720	Human lat
25	16	4.5	3382	6	ABZ35336	Human gen
26	13	3.6	307	7	ADA05695	Human NOV
27	13	3.6	358	7	ADA05693	Human NOV
28	13	3.6	436	7	ADA05699	Human NOV
29	13	3.6	445	7	ADA05691	Human NOV
30	13	3.6	484	7	ADA05687	Human NOV
31	13	3.6	506	6	ABK30302	Human G-p
32	13	3.6	711	7	ADA05020	Human pro
33	13	3.6	783	9	ADD09102	Human kal
34	13	3.6	792	2	AAH80905	Human cDN
35	13	3.6	809	5	AAH87560	DNA encod
36	13	3.6	831	5	AAH85946	DNA encod
37	13	3.6	845	7	ADA05701	Human NOV
38	13	3.6	868	3	AAZ44182	Human neu
39	13	3.6	905	2	AAZ11029	Human ser
40	13	3.6	942	2	AAZ11030	Human ser
41	13	3.6	944	2	AAV84052	Nucleic a
42	13	3.6	988	6	ABZ35510	Human gen
43	13	3.6	994	2	AAH87155	Human pro
44	13	3.6	994	6	ABZ35501	Human gen
45	13	3.6	999	2	AAZ34225	Human PRO

ALIGNMENTS

RESULT 1
ABQ93901
ID ABQ93901 standard; DNA; 1102 BP.
XX AC ABQ93901;
XX AC ABQ93901;
DT 01-NOV-2002 (first entry)
DE Human prostatin precursor-like NOV14a DNA, SEQ ID NO:43.
KW Human; NOVX; neurological disorder; Alzheimer's disease;
KW Huntington's disease; Parkinson's disease; pain; behavioural disorder;
KW addition; tubercous sclerosis; cancer; immune disorder; allergy;
KW autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;
KW thyroiditis; cardiovascular disease; hypertension; reproductive disorder;
KW endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;
KW pancreatic cirrhosis; glomerular endotheliosis; bacterial infection;
KW polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
KW atherosclerosis; cell signal processing-related disorder;
KW metabolic pathway regulation disorder; cytostatic; neuroprotective;
KW antiinflammatory; immunosuppressive; analgesic; antithrombotic;
KW dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;
KW differentiation; proliferation; motility; haematopoiesis; wound healing;
KW angiogenesis; forensic biology; transgenic animal; drug screening;
KW gene therapy; NOV14a; prostatin precursor-like; chromosome 16; gene; ds.

Homo sapiens.

Location/Qualifiers
1..18
/*tag= a
19..1050
CDS

Db 919 ACCAGCCAGAGACCCAGTCAGATTGTTTACATCAAGGCACTTCCTGGATTCTGCC 978
QY 321 ArgileLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrIleSer 340
Db 979 AGAATCCTTTTGAGGCCCTGTCTCCCATATATCATCATAGTCTCAACTGGGACCAAAAGC 1038
QY 341 LeuValLeuProTrpLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
Db 1039 CTTGTCTCTCCCTGGCTCTCTCCACACTCTCTCTGGGCTCTGGGGGTTC 1089
RESULT 2
ID ABQ93902 standard; DNA; 1102 BP.
XX AC ABQ93902;
XX 01-NOV-2002 (first entry)
XX Human prostatin precursor-like NOV14b DNA, SEQ ID NO:45.
XX Human; NOVX; neurological disorder; Alzheimer's disease;
KW Huntington's disease; Parkinson's disease; pain; behavioural disorder;
KW addition; tuberous sclerosis; cancer; immune disorder; allergy;
KW autoimmune disease; myasthenia gravis; asthma; arthritis; diabetes;
KW thyroiditis; cardiovascular disease; hypertension; reproductive disorder;
KW endometriosis; incontinence; psoriasis; scleroderma; alopecia; ulcer;
KW pancreatic disease; glomerular endotheliosis; bacterial infection;
KW polycystic kidney disease; endocrine disorder; obesity; cardiomyopathy;
KW atherosclerosis; cell signal processing-related disorder;
KW metabolic pathway regulation disorder; cytostatic; neuroprotective;
KW antiinflammatory; immunosuppressive; analgesic; antiatherosclerotic;
KW dermatological; antibacterial; antiarthritic; hepatotropic; neurogenesis;
KW differentiation; proliferation; motility; haematopoiesis; wound healing;
KW angiogenesis; forensic biology; transgenic animal; drug screening;
KW gene therapy; NOV14b; prostatin precursor-like; gene; ds.
XX Homo sapiens.
XX Location/Qualifiers
FH 1. .18
FT 5'UTR /*tag= a
FT 19. .1092
FT /*tag= b
FT /*product= "NOV14b"
FT 1093. .1102
FT /*tag= c
XX WO200253742-A2.
XX 11-JUL-2002.
XX 07-JAN-2002; 2002WO-US000375.
XX 05-JAN-2001; 2001US-0260018P.
XX 08-JAN-2001; 2001US-0260360P.
XX 28-FEB-2001; 2001US-0272411P.
XX 02-MAR-2001; 2001US-0272817P.
XX 05-JUL-2001; 2001US-0303231P.
XX 12-JUL-2001; 2001US-0305060P.
XX 10-SEP-2001; 2001US-0318405P.
XX 12-SEP-2001; 2001US-0318700P.
XX 04-JAN-2002; 2002US-00037417.
XX (CURA-) CURAGEN CORP.
XX Kekuda R, Alsbrook JP, Tchernev VT, Liu X, Spytek KA;
PI Pattarajan M, Grosse WM, Lepley DM, Burgess CE, Vernet CM, Li L;
PI Gorman L, Edinger S, Sciore P, Ellerman X, Malyankar U;
PI Rothenberg M, Stone D, Boldog F, Guo X, Shenoy S, Anderson D;
PI Padigaru M, Taupier RU, Miller CE, Eisen A;
XX WPI; 2002-583619/62.

DR P-PSDB; AB09524.
XX Novel polypeptides and nucleic acids homologous to transmembrane
PT receptor, thymosin, neuromodulin-like family of proteins for diagnosing,
PT treating cancer, atherosclerosis, neurological, skin and autoimmune
XX disorders.
PS Claim 9a; Page 143; 323pp; English.
XX The invention relates to 24 novel human proteins designated NOV1-NOV14
CC (AB09501-AB09524), collectively referred to as NOVX proteins, and
CC nucleic acids encoding them (AB093879-AB093902). NOVX proteins and
CC nucleotides are useful in the treatment, diagnosis or prevention of NOVX-
CC associated disorders or in the manufacture of a medicament for treating
CC such disorders, with specific applications described for each of the 24
CC NOVX proteins, based on their homology to known proteins. Various
CC disorders are associated with NOVX proteins including neurological
CC disorders (e.g., Alzheimer's, Huntington's and Parkinson's diseases),
CC pain, behavioural disorders, addiction, tuberculous sclerosis, cancers
CC (e.g., colorectal cancer, leukaemia and osteosarcoma), immune disorders
CC (e.g., allergies and autoimmune diseases), myasthenia gravis, asthma,
CC various forms of arthritis, diabetes, thyroiditis, cardiovascular disease
CC (e.g., hypertension), reproductive disorders, endometriosis,
CC incontinence, psoriasis, scleroderma, alopecia, ulcers, pancreatitis,
CC cirrhosis, glomerular endotheliosis, polycystic kidney disease, endocrine
CC disorders, obesity, bacterial infections and particularly cardiomyopathy,
CC atherosclerosis, cell signal processing-related disorders and disorders
CC of metabolic pathway regulation. NOVX nucleic acids and polypeptides may
CC be used to identify cellular receptors or downstream effectors which
CC binds to a NOVX protein, and are also useful as targets for the
CC identification of small molecules that modulate or inhibit processes such
CC as neurogenesis, cell differentiation, cell motility, cellular
CC proliferation, haematopoiesis, wound healing and angiogenesis. NOVX
CC nucleic acid sequences can be used to identify a cell or tissue type and
CC are useful as a source of primers or probes for forensic biology and for
CC identifying and cloning NOVX homologues in other cell types. Cells
CC comprising NOVX nucleic acids are useful for producing non-human
CC transgenic animals which are useful for studying the function and
CC activity of NOVX proteins and for identifying and evaluating modulators
CC of NOVX activity. The present sequence represents DNA encoding the
CC prostatin precursor-like protein NOV14b
XX SQ Sequence 1102 BP; 162 A; 371 C; 353 G; 216 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0 Length: 1102
Score: 357.00 Matches: 357
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-037-417-46 (1-357) x ABQ93902 (1-1102)
QY 1 MetaLaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAlaAsnSerAspSer 20
Db 19 ATGGCCAGAGGGGGTCTGGGGCTTGGGAGCTGGGGGCTGTGGCCAAATTCGACTCA 78
QY 21 TyrSerLeuTyrGlyLeuValProSerGlyProAlaArgGlyProProTyrCysGlyArg 40
Db 79 TACTCACTTTACGGGTGGTGGTCCGACCGCTAGGGGGCCCCCGTACTGGGGGCGC 138
QY 41 ProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrp 60
Db 139 CCGAGCCCTCGCCCGCATCTGGGGGGGCTCAACCGCCAGCCGGGACCTTGGCTTGG 198
QY 61 GlnValSerLeuHisHisGlyGlyGlyHisIleCysGlyGlySerLeuLeuAlaProSer 80
Db 199 CAGTGAGCTCGCACCATCGAGGTGGCCACATCTGGGGGGGCTCCCTCATCGCCCTCC 258
QY 81 TrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluProAlaAlaGlu 100
Db 259 TGGGTCTCTCTCGGCTGCTCACTGTCTTTCATGACGAATGGGACGTTGGAGCCCGGCGCAG 318

Thu Feb 26 13:50:24 2004

us-10-037-417-46.olip2n.rng

```
QY 101 TrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThrArg 120
Db 319 TGGTCGGTACTGCTGGCGGTGCTACCTCCAGACGGGGCCCTGGACGGCGGCACACCCGC 378
QY 121 AlaValAlaIleValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
Db 379 GCAGTGGCGCGCATCTGTGGTGGCGGCCNACTACAGCCAGTGGAGCTGGGCGCGACCTG 438
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrrProValCysLeu 160
Db 439 GCCCTGCTGGCGCTGCTCCAGCCAGCTGGCGCCCGCGCTGTGGCCCTGTCTGGCTG 498
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrrAlaThrGlyTrrGlyAsp 180
Db 499 CCCCCTGGCTCACACCGCTTGTGTGACGGCACCGCTGTGGGCCACCGCTGGTGGAGAC 558
QY 181 ValGlnGluAlaAspProLeuProLeuProTrrValLeuGlnGluValGluLeuArgLeu 200
Db 559 GTCCAGAGGGAGATCCCTGCTCTCCCTGGGTGCTACAGGAAGTGGAGCTAAGGCTG 618
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
Db 619 CTGGCGAGGCCACCTGTCAATGTCTTACAGCCAGCCCGGTCCCTTCAACCTCACTCTC 678
QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTrrProGluGlyArgArgAspThrCysGln 240
Db 679 CAGATATTGCCAGGGAGTGTGTGTGTGCTACCCAGAGGGCGCGAGCACCTGCCAG 738
QY 241 GlyAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrrPheGlnAlaGlyIle 260
Db 739 GGTGACTCTGGGGGGCCCTGGTCTGTGAGGAAGCGCGCGCTGGTTCAGAGCGGAATC 798
QY 261 ThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThrAlaValAla 280
Db 799 ACCAGCTTTGGTGTGGTGTGGACGAGAAACCGCCCTGGAGTTTTCATGCTGTGGCT 858
QY 281 ThrTyrGluAlaTrrPheArgGluGlnValMetGlySerGluProGlyProAlaPhePro 300
Db 859 ACCTATGAGGCATGGATACGGGAGCAGGTGATGGGTTCAGAGCTGGGGCTGCTTCCC 918
QY 301 ThrGlnProGlnTrrGlnSerAspCysLeuHisGlnThrAlaPheLeuAspSerAla 320
Db 919 ACCAGCCCCAGAGACCCAGTCAGATTGTATCATCAACAGCGCATTCCTGGATTCTGCC 978
QY 321 ArgIleLeuLeuArgProLeuSerHisIleSerValGlyValSerThrGlyThrLysSer 340
Db 979 AGAATCCTTTTGGAGCCCTTGTCCCATATATCATAGTAGAGTCTCACTGGGACCAAGC 1038
QY 341 LeuValLeuProTrrPheLeuSerProHisSerLeuLeuGlyLeuTrrGlyPhe 357
Db 1039 CTGTGCTCTCCCTGGCTCTCTCCACACTCTCTCTGGGGCTCTGGGGGTTT 1089
RESULT 3
ID ABK31795
XX ABK31795 standard; DNA; 2457 BP.
XX
AC ABK31795;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA encoding novel human protease #52.
XX
KW Human; protease; cancer; immune-related disorder; cardiovascular disease;
KW neuronal-associated disease; metabolic disorder; inflammatory disorder;
KW nervous system disorder; sexual dysfunction; pain; mood disorder;
KW hypertension; psychotic disorder; neurological disorder; dyskinesia;
KW viral infection; human immunodeficiency virus; HIV; non-viral infection;
KW ocular disease; cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
PN W0200200860-A2.
```

```
XX 03-JAN-2002.
XX
XX 26-JUN-2001; 2001WO-US020171.
XX
XX 26-JUN-2000; 2000US-0214047P.
XX
XX (SUGE-) SUGEN INC.
XX
XX Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
XX Charyczak G;
XX
XX WPI; 2002-139913/18.
XX
XX P-PSDB; AAU82753.
XX
XX Nucleic acids encoding novel human proteases, useful for useful for
XX treating diseases and disorders such as cancers, immune-related diseases
XX and disorders, cardiovascular disease (e.g. restenosis) and inflammatory
XX disorders.
XX
XX Claim 26; Fig 1S5-TR; 313pp; English.
XX
XX The present invention relates to the isolation of novel human proteases,
XX and the nucleic acids encoding them. The sequences of the invention are
XX useful for treating diseases and disorders such as cancers (e.g. breast,
XX colon, lung), immune-related diseases and disorders (e.g. inflammatory
XX diseases and asthma), cardiovascular diseases (e.g. restenosis and
XX coronary thrombosis), brain or neuronal-associated diseases, metabolic
XX disorders (e.g. diabetes, obesity), inflammatory disorders (e.g.
XX rheumatoid arthritis and psoriasis), central or peripheral nervous system
XX diseases, migraines, pain, sexual dysfunction, mood disorders, attention
XX disorders, neurological disorders, hypotension, hypertension, psychotic
XX disease) and dyskinesias. The nucleic acids and polypeptides are also
XX useful for treating viral infections caused by human immunodeficiency
XX virus (HIV), and non-viral infections such as ocular disease (e.g.
XX glaucoma) and macular degeneration. ABK31744-ABK31802 represent DNA
XX sequences encoding for the novel human proteases of the invention
XX
XX Sequence 2457 BP; 377 A; 851 C; 797 G; 432 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 7,48e-240 Length: 2457
Score: 272.00 Matches: 272
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.19% Indels: 0
DB: Gaps: 6
US-10-037-417-46 (1-357) x ABK31795 (1-2457)
QY 38 CysGlyArgProGluProSerAlaArgIleValGlyGlySerAsnAlaGlnProGlyThr 57
Db 112 TGGCGGGCGCCCTGAGCCCTCGGCCCGCATCTGGGGGGGCTCAAAACGGCGACCGGGCACC 171
QY 58 TrrProTrrGlnValSerLeuHisGlyGlyGlyHisIleCysGlyGlySerLeu 77
Db 172 TGGCTTGGCAAGTGGAGCCCTGACCATGGAGGTGGCCACATCTCGGGGGCTCCCTCATC 231
QY 78 AlaProSerTrrValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
Db 232 GCCCCTCTGGGTCTCTCCGCTGCTCACTTTTCATGACGAATGGAGCGCTGGAGCCC 291
QY 98 AlaAlaGluTrrSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
Db 292 GCGGCCGAGTGGTGGTACTCTGGGGCTGCACTCCAGGACGGGGCCCTTGGACGGCGCG 351
QY 118 HisThrArgAlaValAlaAlaIleValValProAlaAsnTrrSerGlnValGluLeuGly 137
Db 352 CACACCCGCGAGTGGCGGCATCTGTGTGCCGCCAACTACAGCAAGTGGAGCTGGGC 411
QY 138 AlaAspLeuAlaLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrrPro 157
```


QY 49 GlyGlySerAsnAlaGlnProGlyThrTrpProTropGlnValSerLeuHisHisGlyGly 68
 DB 177 GGGGGCTCAACCGGCGAGCGGCACTGGACCTTGGCACTGAGCTGCACCATGGAGGT 236
 QY 69 GlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCys 88
 DB 237 GGGCACATCTGGGGGCTCCCTCATCGCCCTCTCTGGGTCTCTCCGCTGCTCATGT 296
 QY 89 PheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHis 108
 DB 297 TTATGACGAATGGAGCCCTGGAGCCCGCGCCGAGTGGTGGTGGGGCTGCAC 356
 QY 109 SerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaIleValValPro 128
 DB 357 TCCAGGACGGGGCTGGAGCGGCGCACACCGCGCAGTGGCGCCATCGTGGTGGCG 416
 QY 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerPro 148
 DB 417 GCCAACTACAGCCAAAGTGGAGCTGGGGCGCGACCTGGCCCTGTGGCTGCACCC 476
 QY 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168
 DB 477 GCACAGCTGGGGCGGGCGGCGCGCACACCGCGCAGTGGCGCCATCGTGGTGGCG 536
 QY 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
 DB 537 CACGGCACCGCTGCTGGGCCACCGGCTGGGGAGAC 572
 RESULT 9
 AAK93456
 ID AAK93456 standard; cDNA; 670 BP.
 XX AC AAK93456;
 XX DT 06-NOV-2001 (first entry)
 XX DE Human cDNA clone representative sequence, SEQ ID NO: 1916.
 XX KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX OS Homo sapiens.
 XX PN EP1130094-A2.
 XX PD 05-SEP-2001.
 XX PF 07-JUL-2000; 2000EP-00114089.
 XX PR 08-JUL-1999; 99JP-00194486.
 XX PR 11-JAN-2000; 2000JP-00118774.
 XX PR 02-MAY-2000; 2000JP-00183765.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX DR WPI; 2001-524255/58.
 XX PT 830 Primers useful for synthesizing full length cDNA clones and their use
 XX PT in genetic manipulation.
 XX PS Example 11; SEQ ID NO 1916; 1380pp + Sequence Listing; English.
 XX CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesizing the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special

CC methods. The present sequence was used as the representative sequence
 CC from a human clone which was used in homology searches to identify the
 CC clone. Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in CD-ROM format directly from
 CC EPO
 XX SQ Sequence 670 BP; 97 A; 245 C; 205 G; 118 T; 0 U; 5 Other;
 Alignment Scores:
 Pred. No.: 1e-111 Length: 670
 Score: 132.00 Matches: 132
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.97% Indels: 0
 DB: 4 Gaps: 0
 US-10-037-417-46 (1-357) x AAK93456 (1-670)
 QY 49 GlyGlySerAsnAlaGlnProGlyThrTrpProTropGlnValSerLeuHisHisGlyGly 68
 DB 177 GGGGGCTCAACCGGCGAGCGGCACTGGACCTTGGCACTGAGCTGCACCATGGAGGT 236
 QY 69 GlyHisIleCysGlyGlySerLeuIleAlaProSerTrpValLeuSerAlaAlaHisCys 88
 DB 237 GGGCACATCTGGGGGCTCCCTCATCGCCCTCTCTGGGTCTCTCCGCTGCTCATGT 296
 QY 89 PheMetThrAsnGlyThrLeuGluProAlaAlaGluTrpSerValLeuLeuGlyValHis 108
 DB 297 TTATGACGAATGGAGCCCTGGAGCCCGCGCCGAGTGGTGGTGGGGCTGCAC 356
 QY 109 SerGlnAspGlyProLeuAspGlyAlaHisThrArgAlaValAlaIleValValPro 128
 DB 357 TCCAGGACGGGGCTGGAGCGGCGCACACCGCGCAGTGGCGCCATCGTGGTGGCG 416
 QY 129 AlaAsnTyrSerGlnValGluLeuGlyAlaAspLeuAlaLeuLeuArgLeuAlaSerPro 148
 DB 417 GCCAACTACAGCCAAAGTGGAGCTGGGGCGCGACCTGGCCCTGTGGCTGCACCC 476
 QY 149 AlaSerLeuGlyProAlaValTrpProValCysLeuProArgAlaSerHisArgPheVal 168
 DB 477 GCACAGCTGGGGCGGGCGGCGCGCACACCGCGCAGTGGCGCCATCGTGGTGGCG 536
 QY 169 HisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
 DB 537 CACGGCACCGCTGCTGGGCCACCGGCTGGGGAGAC 572
 RESULT 10
 ABK13566/c
 ID ABK13566 standard; cDNA; 456 BP.
 XX AC ABK13566;
 XX DT 08-MAY-2002 (first entry)
 XX DE Human prostatic-like serine protease cDNA #2.
 XX KW Human; prostatic-like serine protease; cytosolic; antiatherosclerotic;
 KW virucidin; osteopathic; antiinflammatory; vasotropic; neuroprotective;
 KW trypsin-like; metastasis; autoimmune lesion; atherosclerosis; gene;
 KW renal failure; malignant cell; tumour angiogenesis; enzyme; inflammation;
 KW chronic obstructive pulmonary disease; COPD; restenosis; atherosclerosis;
 KW neurodegenerative disease; prion protein; infection; amyloid plaque;
 KW Gensmann-Strausler Syndrome; viral infection; Scrapie;
 KW Creutzfeldt-Jakob disease; metastatic cancer; lipid accumulation;
 KW osteoporosis; Paget's disease; ss; EST; expressed sequence tag.
 XX OS Homo sapiens.
 XX PN WO200198467-A2.
 XX PD 27-DEC-2001.
 XX PF 22-JUN-2001; 2001WO-EP007117.

XX 23-JUN-2000; 2000US-0213588P.
PR 20-MAR-2001; 2001US-0276909P.
XX (FARB) BAYER AG.
FA Xiao Y, Morozov V;
XX WPI; 2002-114576/15.
XX
XX Novel human prostatic-like serine protease polypeptide and polynucleotide
PT which can be regulated for treating metastasis of malignant cells,
PT inflammation, atherosclerosis, neurodegenerative disease and infections.
XX
XX Disclosure; Fig 3; 11pp; English.
XX
XX This invention comprises the cDNA and protein sequences of an isolated
CC prostatic-like serine protease and reagents and methods for regulating
CC the human prostatic-like enzyme activity. Prostatic is a trypsin-like
CC serine protease purified from human seminal fluid. An antibody specific
CC for prostatic-like serine protease is useful for immunodetection and
CC diagnosis of micro-metastases, autoimmune lesions and renal failure in
CC biopsy specimens, plasma samples and body fluids. The antibody may be
CC used to modulate enzyme activity in a disease, such as metastasis of
CC malignant cells, tumour angiogenesis, inflammation, chronic obstructive
CC pulmonary disease (COPD), atherosclerosis, neurodegenerative disease
CC (e.g. prion protein amyloid plaques of Genstamm-Straussler Syndrome,
CC Creutzfeldt-Jakob disease, Scrapie) or infection, particularly viral
CC infection. The human prostatic-like serine protease gene provides a
CC therapeutic target of decreasing human prostatic-like serine protease
CC activity, in particular for treating or preventing metastatic cancer. The
CC agonists and antagonists of the nucleotide sequence may be used to mimic,
CC augment and inhibit the enzyme activity which may be useful to treat
CC osteoporosis, Paget's disease and degradation of bone implants
CC particularly dental implants. Altered levels of human prostatic-like
CC serine protease activity inhibits both smooth muscle cell proliferation
CC and lipid accumulation and inhibit the progression of restenosis and
CC atherosclerosis. The nucleic acid sequence is also useful in diagnostic
CC assays for detecting diseases and abnormalities or susceptibility to
CC diseases related to the presence of mutations in nucleic acid sequences
CC which encode the enzyme. The present sequence represents the human
CC prostatic-like serine protease #2 nucleotide sequence of the invention
XX
SQ Sequence 456 BP; 74 A; 151 C; 167 G; 62 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.: 3,378-108 Length: 456
Score: 128.00 Matches: 128
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.85% Indels: 0
DB: 6 Gaps: 0

US-10-037-417-46 (1-357) x ABK13566 (1-456)

QY 121 AlaValAlaAlaIleValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
Db 384 CGAGTGGCGCCATCGTGGTCCGGCAACTACAGCAAGTGGAGTGGCGCGCACCTG 325
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCysLeu 160
Db 324 GCCTCTGGCTGGCTGGCTACCCCGCAGCCTGGGCGCGGTGGCTGTCTGCTG 265
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
Db 264 CCCCGGCTCACACCGCTTCGTGCACGCGACCGCTGTGGCCACCGCTGGCGAGAC 205
QY 181 ValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGluLeuArgLeu 200
Db 204 GTCCAGGAGGAGATCTCTGCTCTCTCCCTGGGTGCTACAGGAAGTGGAGCTAAGGCTG 145
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220

Db 144 CTGGCGGAGGCCACCTGTCAATGTCTCTACAGCAGCCCGGTCCCTTCAACCTCACTCTC 85
QY 221 GlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCysGln 240
Db 84 CAGATATTGCGAGGGATGCTGTGTGGTGTACCCAGAGGGCGCGAGACACCTGCCAG 25
QY 241 GlyAspSerGlyGlyProLeuVal 248
Db 24 GGTGACTCTGGGGGCGCCCTGGTC 1

RESULT 11

ID AAS71079 standard; cDNA; 2298 BP.
XX AAS71079;
AC AAS71079;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #6883.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
(HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ASG06892.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

Claim 1; SEQ ID NO 6883; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 2298 BP; 468 A; 774 C; 706 G; 350 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 9.73e-38 Length: 2298
 Score: 52.00 Matches: 52
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.57% Indels: 0
 DB: 5 Gaps: 0

US-10-037-417-46 (1-357) x AAS71079 (1-2298)

QY 258 AlaGlyIleThrSerPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
 DB 424 GCAGAAATCACAGCTTGGCTTTGGCTGTGGACGGAGAAACCGCCCGGAGTTTCACT 483
 QY 278 AlaValAlaThrTyrGluAlaTrpIleArgGluInValMetGlySerGluProGlyPro 297
 DB 484 GCTGTGGCTACTATGAGGCATGTATACGGGACGAGTGATGGTTCAGAGCCTGGGCT 543
 QY 298 AlaPheProThrGlnProGlnIleThrGlnSerAsp 309
 DB 544 GCCTTTCCACCCAGCCCGACAGACCCAGTCAGAT 579

RESULT 12

AAH78206

ID AAH78206 standard; DNA; 768 BP.

XX AC

AAH78206;

XX AC

XX 26-NOV-2001 (first entry)

XX DE

XX Nucleotide sequence of a human secreted polypeptide.

KW Human; secreted polypeptide; nervous disease; muscular disease; tumour;
 KW Gastrointestinal ulceration; spinal cord disease; trachea disease;
 KW thyroid gland disease; ovary disease; prostate disease; heart disease;
 KW renal gland disease; small intestine disease; thymus disease;
 KW lymph node disease; muscular system disease; colon disease;
 KW lipase deficiency; cystic fibrosis; pancreatitis; clot formation;
 KW myocardial infarction; angioplasty; liver disease; coagulation disorder;
 KW microbial disease; immune disorder; inflammation; transplant rejection;
 KW bone thickness; bone density; ferroxidase loss; apoptosis;
 KW vascular smooth cell proliferation; vaccine; ss.

XX OS Homo sapiens.

XX FH

XX Key Location/Qualifiers

XX CDS

XX 1..768

XX /*tag= a

XX /product= "secreted polypeptide"

XX FT

XX WO200166690-A2.

XX PN

XX 13-SEP-2001.

XX PD

XX 05-MAR-2001; 2001WO-US007143.

XX PF

XX 06-MAR-2000; 2000US-0187107P.

XX PR

XX 13-MAR-2000; 2000US-0188916P.

XX PR

XX 03-OCT-2000; 2000US-0236874P.

XX PR

XX 03-OCT-2000; 2000US-0237846P.

XX XX

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX PA

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX XX

XX Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;

XX PI

XX WPI: 2001-570768/64.

XX DR

XX P-PSDB; AAG67514.

XX XX

XX Novel isolated secreted polypeptide useful for treating nervous and

XX PT muscular diseases, gastrointestinal ulceration, coagulation and immune

XX PT disorders, microbial diseases, inflammation and transplant rejection.

XX XX

XX Claim 2; Page 44-45; 102pp; English.

XX The present sequence encodes a human secreted polypeptide. The secreted
 CC polypeptides and polynucleotides are useful for treating nervous and
 CC muscular diseases, for inhibiting tumour formation and metastasis, for
 CC treating gastrointestinal ulceration, for preventing and treating
 CC diseases in spinal cord, thyroid gland, ovary, prostate, renal gland,
 CC small intestine, heart, trachea, thymus, lymph node, muscular system and
 CC colon, for treating lipase deficiency in cystic fibrosis and
 CC pancreatitis, for treating undesirable clot formation such as myocardial
 CC infarction, during angioplasty and all surgical procedures that require
 CC decreased blood clot formation, for treating liver diseases, coagulation
 CC disorders and microbial diseases, for treating immune disorders, for
 CC treating inflammation and transplant rejection, for enhancing bone
 CC thickness and increasing bone density, for reducing the loss of essential
 CC ferroxidases, for suppressing apoptosis, and for regulating vascular
 CC smooth cell proliferation. They may also be used as vaccines

XX Sequence 768 BP; 125 A; 280 C; 224 G; 139 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 9.83e-32 Length: 768
 Score: 45.00 Matches: 45
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 12.61% Indels: 0
 DB: 4 Gaps: 0

US-10-037-417-46 (1-357) x AAH78206 (1-768)

QY 47 IleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerIleuHis 66

DB 1 ATCTGGGGGGCTCAAAACGGCGACCGGCGACCTGGCGCTGGCAAGTGAGCCTGCACCAT 60

QY 67 GlyGlyGlyHisIleCysGlyGlySerIleuIleAlaProSerTrpValLeuSerAlaAla 86

DB 61 GGAGTGGCCACATCTCGGGGGCTCCCTCAFGCCCCCTCTGGGHCCTTCCTCGCT 120

QY 87 HisCysPheMetThr 91

DB 121 CACTGTTTCATGAGG 135

RESULT 13

ABA21236/c

ID ABA21236 standard; DNA; 8899 BP.

XX AC

ABA21236;

XX AC

DT 23-JAN-2002 (first entry)

XX DE

XX Human nervous system related polynucleotide SEQ ID NO 13567.

XX KW

XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;

XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;

XX antiparkinsonian; antischlicking; antianaemic; antiarthritis; cancer;

XX antineumatic; hepatotropic; cerebroprotective; antiinflammatory;

XX antiallergic; antidiabetic; antilulcer; anticonvulsant; antifungal;

XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;

XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX OS Homo sapiens.

XX PN

XX WO200159063-A2.

XX XX

XX 16-AUG-2001.

XX XX

XX 17-JAN-2001; 2001WO-US001334.

XX PF

XX 31-JAN-2000; 2000US-0179065P.

XX PR

XX 04-FEB-2000; 2000US-0180628P.

XX PR

XX 24-FEB-2000; 2000US-0184664P.

XX PR

XX 02-MAR-2000; 2000US-0186350P.

XX PR

XX 16-MAR-2000; 2000US-0189874P.

XX PR

XX 17-MAR-2000; 2000US-0190076P.

us-10-037-417-46.olip2n.rng

Thu Feb 26 13:50:24 2004

PR	18-APR-2000;	2000US-0198123P.	PR	20-OCT-2000;	2000US-0240960P.
PR	19-MAY-2000;	2000US-0205515P.	PR	20-OCT-2000;	2000US-0241785P.
PR	07-JUN-2000;	2000US-0209467P.	PR	20-OCT-2000;	2000US-0241786P.
PR	28-JUN-2000;	2000US-0214896P.	PR	20-OCT-2000;	2000US-0241787P.
PR	30-JUN-2000;	2000US-0215135P.	PR	20-OCT-2000;	2000US-0241808P.
PR	07-JUL-2000;	2000US-0216647P.	PR	20-OCT-2000;	2000US-0241809P.
PR	07-JUL-2000;	2000US-0216880P.	PR	20-OCT-2000;	2000US-0241826P.
PR	11-JUN-2000;	2000US-0217487P.	PR	20-OCT-2000;	2000US-0242221P.
PR	11-JUL-2000;	2000US-0217496P.	PR	01-NOV-2000;	2000US-0244617P.
PR	14-JUL-2000;	2000US-0218290P.	PR	08-NOV-2000;	2000US-0246474P.
PR	26-JUL-2000;	2000US-0220963P.	PR	08-NOV-2000;	2000US-0246475P.
PR	26-JUL-2000;	2000US-0220964P.	PR	08-NOV-2000;	2000US-0246476P.
PR	14-AUG-2000;	2000US-0224518P.	PR	08-NOV-2000;	2000US-0246477P.
PR	14-AUG-2000;	2000US-0224519P.	PR	08-NOV-2000;	2000US-0246478P.
PR	14-AUG-2000;	2000US-0225213P.	PR	08-NOV-2000;	2000US-0246523P.
PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225266P.	PR	08-NOV-2000;	2000US-0246525P.
PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225757P.	PR	08-NOV-2000;	2000US-0246609P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225759P.	PR	08-NOV-2000;	2000US-0246611P.
PR	18-AUG-2000;	2000US-0226799P.	PR	08-NOV-2000;	2000US-0246613P.
PR	22-AUG-2000;	2000US-0226681P.	PR	17-NOV-2000;	2000US-0249207P.
PR	22-AUG-2000;	2000US-0226688P.	PR	17-NOV-2000;	2000US-0249208P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249209P.
PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249210P.
PR	30-AUG-2000;	2000US-0228247P.	PR	17-NOV-2000;	2000US-0249211P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249214P.
PR	03-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249215P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249216P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249218P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249219P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0232081P.	PR	17-NOV-2000;	2000US-0249299P.
PR	12-SEP-2000;	2000US-0231968P.	PR	01-DEC-2000;	2000US-0250390P.
PR	14-SEP-2000;	2000US-0232397P.	PR	01-DEC-2000;	2000US-0250391P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251160P.
PR	14-SEP-2000;	2000US-0232399P.	PR	05-DEC-2000;	2000US-0251303P.
PR	14-SEP-2000;	2000US-0232400P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232401P.	PR	06-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251866P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251869P.
PR	21-SEP-2000;	2000US-0234223P.	PR	08-DEC-2000;	2000US-0251989P.
PR	21-SEP-2000;	2000US-0234274P.	PR	11-DEC-2000;	2000US-0251990P.
PR	25-SEP-2000;	2000US-0234597P.	PR	11-DEC-2000;	2000US-0254097P.
PR	25-SEP-2000;	2000US-0234598P.	PR	05-JAN-2001;	2001US-0259678P.
PR	26-SEP-2000;	2000US-0234599P.	XX		
PR	27-SEP-2000;	2000US-0235484P.	XX		
PR	27-SEP-2000;	2000US-0235834P.	XX		
PR	27-SEP-2000;	2000US-0235836P.	XX		
PR	29-SEP-2000;	2000US-0236327P.	XX		
PR	29-SEP-2000;	2000US-0236367P.	XX		
PR	29-SEP-2000;	2000US-0236368P.	XX		
PR	29-SEP-2000;	2000US-0236369P.	XX		
PR	29-SEP-2000;	2000US-0236370P.	XX		
PR	02-OCT-2000;	2000US-0236802P.	XX		
PR	02-OCT-2000;	2000US-0237037P.	XX		
PR	02-OCT-2000;	2000US-0237038P.	XX		
PR	02-OCT-2000;	2000US-0237039P.	XX		
PR	02-OCT-2000;	2000US-0237040P.	XX		
PR	13-OCT-2000;	2000US-0239935P.	XX		
PR	13-OCT-2000;	2000US-0239937P.	XX		

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases.

Disclosure; SEQ ID NO 13567; 1701pp + Sequence Listing; English.

The invention relates to novel genes (AB11004-AB21534) and proteins (AB14678-AB18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are

CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 8899 BP; 1841 A; 2663 C; 2617 G; 1778 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.73e-05 Length: 8899
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.76% Indels: 0
DB: 5 Gaps: 0

US-10-037-417-46 (1-357) x ABA21236 (1-8899)

QY 341 LeuValLeuProTrieSerProHisSerLeuLeuGlyLeuTrieGlyPhe 357
Db 747 CTTCCTCCCTGGCTCTCTCCACACTCTCTCTCTGGGCTCTGGGGTTT 697

RESULT 14
ABN85393
ID ABN85393 standard; DNA; 1161 BP.
AC ABN85393;
XX
DT 21-OCT-2002 (first entry)
XX
DE Human NOV14b, prostasin-like protein, coding sequence.
XX
KW Human; NOV14b; cytostatic; Cardiant; Antiinflammatory; Immunosuppressive;
KW Antiallergic; Haemostatic; Anti-HIV; Antidiabetic; Anorectic;
KW Antiasthmatic; Nephrotropic; Hepatotropic; Neuroprotective; Nootropic;
KW Antibacterial; Virucide; Antiparasitic; Relaxant; Anticonvulsant;
KW Gene Therapy; NOV; cancer; heart disease; inflammation;
KW autoimmune disorder; allergy; blood disorder; AIDS; diabetes; obesity;
KW asthma; IgA nephropathy; cirrhosis; arthritis; Alzheimer's disease;
KW infection; stroke; muscular dystrophy; epilepsy; wasting disorder;
KW prostasin-like protein; chromosome 16; gene; ds.

XX Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 1..1161
FT /*tag= a
FT /product= "NOV14b"
XX
PN WO200255704-A2.
XX
PD 18-JUL-2002.
XX
PF 09-JAN-2002; 2002WO-US000554.
XX
PR 09-JAN-2001; 2001US-0260417P.
PR 10-JAN-2001; 2001US-0260831P.
PR 28-FEB-2001; 2001US-0272338P.
PR 09-MAR-2001; 2001US-0274876P.
PR 18-APR-2001; 2001US-0284704P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Padigaruru M, Li L, Zerhusen BD, Casman SJ, Shenoy S, Spytek KA;

PI Zhong M, Gangolli EA, Burgess CE, Patturajan M, Vernet CAM;
PI Taylor S, Tchernev VT, Miller CE, Guo X, Roldog FL, Grosse WM;
PI Alsbrook JP, Gerlach V, Edinger S, Rothenberg ME, Ellerman K;
PI Macdougall J, Malyankar U, Millet I, Peyman J, Smithson G;
PI Gunther E, Stone DJ;
XX
WPI; 2002-590674/63.
P-PSDB; ABB98416.
XX
NOVX polypeptides and encoding polynucleotides, useful for preventing or
PT treating NOVX-associated disorders e.g. cancer, inflammation, or
PT Alzheimer's disease, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
PS Claim 9; Page 99; 358pp; English.
XX
CC The present sequence is a coding sequence for a NOV protein. The NOV
CC proteins and coding sequences are useful for treating or preventing NOV-
CC associated disorders or in the manufacture of a medicament for treating
CC the disorders, such as cancer, heart disease, inflammation, autoimmune
CC disorders, allergies, blood disorders, AIDS, diabetes, obesity, asthma,
CC IgA nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections
CC (e.g. bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy,
CC and other wasting disorders associated with chronic diseases. NOV14b is a
CC prostasin-like protein, and the NOV14b coding sequence localises to
CC chromosome 16
XX

SQ Sequence 1161 BP; 219 A; 349 C; 364 G; 229 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.3e-05 Length: 1161
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 6 Gaps: 0

US-10-037-417-46 (1-357) x ABN85393 (1-1161)

QY 1 MetAlaGlnGlyGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 1 ATGGCCAGAAAGGGGGTCTCTGGGCTCTGGGAGCTGGGGGCTGTGGCC 48

RESULT 15
ABN85392
ID ABN85392 standard; DNA; 1726 BP.
XX
AC ABN85392;
XX
DT 21-OCT-2002 (first entry)
XX
DE Human NOV14a, prostasin-like protein, coding sequence.
XX
KW Human; NOV14a; cytostatic; Cardiant; Antiinflammatory; Immunosuppressive;
KW Antiallergic; Haemostatic; Anti-HIV; Antidiabetic; Anorectic;
KW Antiasthmatic; Nephrotropic; Hepatotropic; Neuroprotective; Nootropic;
KW Antibacterial; Virucide; Antiparasitic; Relaxant; Anticonvulsant;
KW Gene Therapy; NOV; cancer; heart disease; inflammation;
KW autoimmune disorder; allergy; blood disorder; AIDS; diabetes; obesity;
KW asthma; IgA nephropathy; cirrhosis; arthritis; Alzheimer's disease;
KW infection; stroke; muscular dystrophy; epilepsy; wasting disorder;
KW prostasin-like protein; chromosome 16p11.2; gene; ds.

XX Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 229..1152
FT /*tag= a
FT /product= "NOV14a"
XX
PN WO200255704-A2.
XX
PD 18-JUL-2002.

Thu Feb 26 13:50:24 2004

us-10-037-417-46.olip2n.rng

```
XX 09-JAN-2002; 2002WO-US000554.
XX
XX 09-JAN-2001; 2001US-0260417P.
XX 10-JAN-2001; 2001US-0260831P.
PR 28-FEB-2001; 2001US-0272338P.
PR 09-MAR-2001; 2001US-0274876P.
PR 18-APR-2001; 2001US-0284704P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Padigaru M, Li L, Zerhusen BD, Casman SJ, Shenoy S, Spytek KA;
PI Zhong M, Gangolli EA, Burgess CE, Patturajan M, Vernet CM;
PI Taylor S, Tcherven V, Miller CE, Guo X, Boldog FL, Grose WM;
PI Alsobrook JP, Gerlach V, Edinger S, Rothenberg ME, Ellerman K;
PI Macdougall J, Malyankar U, Millet I, Feyman J, Smithson G;
PI Gunther E, Stone DJ;
XX
XX WPI; 2002-590674/63.
DR P-PSDB; ABB98415.
XX
XX NOVX polypeptides and encoding polynucleotides, useful for preventing or
PT treating NOVX-associated disorders e.g. cancer, inflammation, or
PT Alzheimer's disease, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
XX Claim 9; Page 97-98; 358pp; English.
XX
XX The present sequence is a coding sequence for a NOV protein. The NOV
CC proteins and coding sequences are useful for treating or preventing NOV-
CC associated disorders or in the manufacture of a medicament for treating
CC the disorders, such as cancer, heart disease, inflammation, autoimmune
CC disorders, allergies, blood disorders, AIDS, diabetes, obesity, asthma,
CC IGA nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections
CC (e.g. bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy,
CC and other wasting disorders associated with chronic diseases. NOV14a is a
CC prostaticin-like protein, and the NOV14a coding sequence localises to
CC chromosome 16p11.2
XX
XX Sequence 1726 BP; 284 A; 581 C; 499 G; 362 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 9e-05 Length: 1726
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 6 Gaps: 0
```

```
US-10-037-417-46 (1-357) x ABN85392 (1-1726)
Qy 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 229 ATGGCCCAAGAGGGGGCTCTGGGGCTGGGAGCTGGGGGGCTGTGGCC 276
```

Search completed: February 25, 2004, 22:21:32
Job time : 431 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 18:57:43 ; Search time 2496 Seconds
(without alignments)
4271.153 Million cell updates/sec

Title: US-10-037-417-46
Perfect score: 357
Sequence: 1 MAQKGVLPGLGAVANS.....TKSLVLPWLSLGLWGF 357

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1

Total number of hits satisfying chosen parameters: 55025477

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEP=frame+ p2n.model -DEV=xlp
-Q/cgn2_1/USPTO spool P/US10037417/runat 24022004 101108 18096/app query.fasta_1.519
-DB=EST -QFMT=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417 @C@N 1 1 3549 @runat 24022004 101108 18096 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_estc2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_in:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	154	43.1	537	10	AW450407
C 2	128	35.9	456	9	AI190509
C 3	57	16.0	930	13	EX436299
C 4	35	9.8	296	10	BF561257
C 5	35	9.8	322	9	AA996811
C 6	23	6.4	351	13	BY236668
C 7	23	6.4	365	13	BY235958
C 8	23	6.4	369	13	BY236911
C 9	23	6.4	430	28	AZ252377
C 10	23	6.4	433	10	BB850607
C 11	23	6.4	454	28	AZ875648
C 12	20	5.6	433	10	BB850564
C 13	17	4.8	875	13	BU164841
C 14	17	4.8	907	14	CD251990
C 15	17	4.8	929	13	BU860221
C 16	16	4.5	249	10	BB593614
C 17	16	4.5	294	9	AA300017
C 18	16	4.5	318	12	BM828821
C 19	16	4.5	360	9	AI393077
C 20	16	4.5	366	9	AI623099
C 21	16	4.5	377	9	AI193435
C 22	16	4.5	406	14	CB116948
C 23	16	4.5	409	9	AU298208
C 24	16	4.5	433	10	BB850795
C 25	16	4.5	462	9	AI343968
C 26	16	4.5	472	9	AI761519
C 27	16	4.5	479	9	AU298317
C 28	16	4.5	479	14	CB996389
C 29	16	4.5	492	10	BF063417
C 30	16	4.5	516	12	BM311864
C 31	16	4.5	527	9	AI658798
C 32	16	4.5	527	12	BG482000
C 33	16	4.5	534	9	AI393068
C 34	16	4.5	537	12	BM837284
C 35	16	4.5	537	12	BM837397
C 36	16	4.5	542	10	BE280845
C 37	16	4.5	555	12	BM837589
C 38	16	4.5	564	12	BM750655
C 39	16	4.5	570	12	BG479276
C 40	16	4.5	570	14	CB125988
C 41	16	4.5	575	9	AI207222
C 42	16	4.5	583	12	BG386531
C 43	16	4.5	590	13	BQ292219
C 44	16	4.5	591	9	AU134797
C 45	16	4.5	594	10	BE281529

ALIGNMENTS

RESULT 1
AW450407/c
LOCUS
DEFINITION
IMAGE:2735037 3', mRNA sequence.
ACCESSION
AW450407.1
VERSION
AW450407.1
KEYWORDS
EST.
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 537)
537 bp mRNA linear EST 17-FEB-2000
UI-H-BI3-akn-g-11-0-UI.sl NCI CGAP Sub5 Homo sapiens cDNA clone

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Oligo-dt track not found, Not I site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares lab Clone distribution: NCI-CGAP clone distribution
 information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward
 POLYA=No.

FEATURES
 source
 1. .537
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2735037"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP Subs"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP Subs
 is a subtracted library derived from NCI CGAP Sub4. The
 NCI CGAP Subs library had 3 million recombinants. A
 single-stranded DNA preparation of NCI CGAP Sub4 was used
 as a tracer in a subtractive hybridization with a driver
 comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM
 3334-3337, 3682-3683, 3798-3803) (IMAGE Clones)
 1322376-1323911, 1456008-1456775, 1500552-1502855);
 NCI CGAP Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
 (IMAGE Clones) 1323912-1325831, 1471368-1472903,
 1492104-1493255); NCI CGAP Lu5 pool 1 LLAM 3575-3582,
 3851-3854 (IMAGE Clones)
 1414920-1417991, 1520904-1522439); NCI CGAP GC4 pool 1 LLAM
 3164-3167, 3716-3720, 3733-3735 (IMAGE Clones)
 1257096-1258631, 1469064-1470983, 1475592-1476743);
 NCI CGAP Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
 (IMAGE Clones) 985608-986759, 1101192-1101959,
 1217928-1220615); NCI CGAP Co10 pool 1 LLAM 2644-2653,
 2871-2872 (IMAGE Clones)
 1057416-1061255, 1144584-1145351). (10% of the driver
 population), plus a pool of 3,840 arrayed clones from
 NCI CGAP Sub1 (IMAGE Clones) 2708616-2710535) and
 NCI CGAP Sub2 (IMAGE Clones) 2710536-2712455) (10% of
 the driver population), plus a pool of 11,136 clones from
 NCI CGAP Sub3 (IMAGE Clones) 2712456-2723591) (10% of the
 driver population), plus a pool of 5,472 clones from
 NCI CGAP Sub4 (IMAGE Clones) 2723592-2728969) (70% of the
 driver population). Subtraction was performed as
 previously described (Bonaldi, Lennon & Soares (1996):
 Normalization and Subtraction: Two Approaches To
 Facilitate Gene Discovery. Genome Research 6, 791-806.
 TAG TISSUE=colon
 TAG LIB=NCI CGAP_Co10
 TAG_SEQ=AAACG"

ORIGIN
 Alignment Scores:
 Pred. No.: 8.71e-127 Length: 537
 Score: 154.00 Matches: 154
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 43.14% Indels: 0
 DB: 10 Gaps: 0

US-10-037-417-46 (1-357) x AW450407 (1-537)

QY 100 GluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAlaHisThr 119
 Db 464 GAGTGGTGGTACTGCTGGCGGTGCACTCCAGGACGGCCCTGGACGGCGCACACC 405
 QY 120 ArgAlaValAlaAlaIleValProAlaAsnTyrSerGlnValGluLeuGlyAlaAsp 139

Db 404 CCGCAGTGGCGCCGATCGTGTGCGGCCCACTACAGCCAGTGGAGTGGCGCGCCGAC 345
 QY 140 LeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpProValCys 159
 Db 344 CTGGCCCTGCTGGCGCTGGCGCTCACCCGCGAGCTGGGCCCCCGCGTGTGGCTGTCTGC 285
 QY 160 LeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGly 179
 Db 284 CTGGCCCGCGCGCTCACCGCTTGTGTGACGCGCACCGCTGTGGGCCACCGGCTGGGGA 225
 QY 180 AspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGluLeuArg 199
 Db 224 GACGTCCAGAGGAGGATCTCTGCTCTCCCTGGGTGTACAGGAAGTGGAGTAAGG 165
 QY 200 LeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThr 219
 Db 164 CTGCTGGCGAGGCGCACCTGTCAATGTCTTACAGCCAGCCGCGTCCCTTCACTCACT 105
 QY 220 LeuGlnIleLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAspThrCys 239
 Db 104 CTCAGATATTGCCAGGGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 45
 QY 240 GlnGlyAspSerGlyGlyProLeuValCysGlnGluGlyGly 253
 Db 44 CAGGCTGACTCTGGGGGGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3

RESULT 2
 AL190509/c
 LOCUS AL190509.1
 DEFINITION Q449f07.x1 Soares fetal heart NBHH19W Homo sapiens cDNA clone
 IMAGE:1732837 3' similar to SW:MCP6_MOUSE P21845 MAST CELL PROTEASE
 6 PRECURSOR ; mRNA sequence.
 AL190509
 VERSION AL190509.1 GI:3741718
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 456)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 712 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 452.
 Location/Qualifiers
 1. .456
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1732837"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares fetal heart NBHH19W"
 /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 GTTACCAATCTGAAGTGGAGCGCGCATCTTTTTTTTTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by
 M.Fatima Bonaldi. This library was constructed from the
 same fetus as the fetal lung library, Soares fetal lung

```

ORIGIN
NbHL19W."

Alignment Scores:
Pred. No.: 1.06e-103 Length: 456
Score: 128.00 Matches: 128
Percent Similarity: 100.00% Conservaive: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.95% Indels: 0
DB: 9 Gaps: 0

US-10-037-417-46 (1-357) x A1190509 (1-456)
QY 121 AlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuGlyAlaAspLeu 140
Db 384 GCAGTGGCCGCCATCTGGTGGTCCCGCCCACTACAGCAAGTGGAGCTGGGGCGGACCTG 325
QY 141 AlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTyrProValCysLeu 160
Db 324 GCCCTGCTGCCCTGGCTCACCAGCCAGCTGGGCCCGCGGTGGCTGTCTGCTG 265
QY 161 ProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGlyTrpGlyAsp 180
Db 264 CCCCGGCGCTTCACACCGCTTCGTGCAGCGCACCGCTGCTGGCCACCGGCTGGGAGAC 205
QY 181 ValGlnGluAlaAspProLeuProLeuProTyrValLeuGlnGluValGluLeuArgLeu 200
Db 204 GTCCAGAGGACGATCTCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 145
QY 201 LeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsnLeuThrLeu 220
Db 144 CTGGCGAGGCCACCTGTCAATGCTCTACAGCAGCCCGGTCCTTCAACCTCACTCTC 85
QY 221 GlnIleuProGlyMetLeuCysAlaGlyTyrProGluGlyArgAspThrCysGln 240
Db 84 CAGATATTCCAGGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25
QY 241 GlyAspSerGlyGlyProLeuVal 248
Db 24 GTGACTCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1

RESULT 3
BX436299
LOCUS
DEFINITION
3-PRIME mRNA sequence.
ACCESSION
BX436299
VERSION
BX436299.1 GI:30787522
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7995.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0CAP001AD09FM1&cluster=7995.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0CAP001AD09FM1.
FEATURES
Location/Qualifiers
1..930
/organism="Homo sapiens"
/mol_type="mRNA"

ORIGIN
/db_xref="taxon:9606"
/clone="CS0CAP001YGI7"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: PCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the PCMVSPORT 6 vector.
Library was not normalized."

Alignment Scores:
Pred. No.: 4.21e-40 Length: 930
Score: 57.00 Matches: 57
Percent Similarity: 100.00% Conservaive: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.97% Indels: 0
DB: 13 Gaps: 0

US-10-037-417-46 (1-357) x BX436299 (1-930)
QY 253 GlyArgTrpPheGlnAlaGlyIleThrSerPheGlyPheGlyCysGlyArgAsnArg 272
Db 70 GCGCGTGGTTCAGGAGGAAATCACAGCTTTGGCTTTGGCTTGACGAGAAACCGC 129
QY 273 ProGlyValPheThrAlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGly 292
Db 130 CTGGAGTTTTCAGTCTGTGCTACCTATGAGGCAATGATACGGAGCAGGTGATGGGT 189
QY 293 SerGluProGlyProAlaPheProThrGlnProGlnLysThrGlnSerAsp 309
Db 190 TCAGAGCTGGGCTGCTTCCACCCAGCCAGAGACCCAGTCAGAT 240

RESULT 4
BF561257/c
LOCUS
DEFINITION
UI-R-CO-hh-h-07-0-UI.r1 UI-R-CO Rattus norvegicus cDNA clone
ACCESSION
BF561257
VERSION
BF561257.1 GI:11670987
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 296)
Ronald,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LNL (info@image.llnl.gov). IMAGE ID= 1781449
Seq primer: M13 Forward.
Location/Qualifiers
1..296
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CO-hh-h-07-0-UI"
/dev_stage="adult"

FEATURES
source

```

/lab host="DH10B (Life Technologies)"
 /clone lib="UI-R-CO"
 /note=Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CO
 library is a subtracted library derived from the UI-R-Al
 and UI-R-E1 libraries. The UI-R-Al library consisted of a
 mixture of individually tagged normalized libraries
 constructed from rat placenta, adult lung, brain, liver,
 kidney, heart, spleen, ovary, and muscle. The UI-R-E1
 library consisted of a mixture of individually tagged
 normalized libraries constructed from 8, 12 and 18-day
 embryo. The tag is a string of 3-5 nucleotides present
 between the Not I site and the oligo-dT track which
 allows identification of the library of origin of a clone
 within the mixture. The subtracted library (UI-R-CO) was
 constructed as follows: PCR amplified cDNA inserts from a
 pool of UI-R-Al and UI-R-E1 clones from which 3' ESTs had
 been derived was used as a driver in a hybridization with
 the pooled UI-R-Al and UI-R-E1 library in the form of
 single-stranded circles. The remaining single-stranded
 circles (subtracted library) was purified by
 hydroxyapatite column chromatography, converted to
 double-stranded circles and electroporated into DH10B
 bacteria (Life Technologies) to generate the UI-R-CO
 library. This procedure has been previously described
 (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
 1996)

Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Oligo-dT track not found, Not I site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.
 Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
 through Research Genetics This clone is also available through the
 I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE
 ID=1781449
 Seq primer: M13 Forward
 POLYA=No.
 Location/Qualifiers
 1..322
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-CO-hh-h-07-0-UI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-CO"
 /note=Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CO
 library is a subtracted library derived from the UI-R-Al
 and UI-R-E1 libraries. The UI-R-Al library consisted of a
 mixture of individually tagged normalized libraries
 constructed from rat placenta, adult lung, brain, liver,
 kidney, heart, spleen, ovary, and muscle. The UI-R-E1
 library consisted of a mixture of individually tagged
 normalized libraries constructed from 8, 12 and 18-day
 embryo. The tag is a string of 3-5 nucleotides present
 between the Not I site and the oligo-dT track which
 allows identification of the library of origin of a clone
 within the mixture. The subtracted library (UI-R-CO) was
 constructed as follows: PCR amplified cDNA inserts from a
 pool of UI-R-Al and UI-R-E1 clones from which 3' ESTs had
 been derived was used as a driver in a hybridization with
 the pooled UI-R-Al and UI-R-E1 library in the form of
 single-stranded circles. The remaining single-stranded
 circles (subtracted library) was purified by
 hydroxyapatite column chromatography, converted to
 double-stranded circles and electroporated into DH10B
 bacteria (Life Technologies) to generate the UI-R-CO
 library. This procedure has been previously described
 (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
 1996)

Alignment Scores: 4.65e-21 Length: 296
 Pred. No.: 35.00 Matches: 35
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 9.80% Gaps: 0
 DB: 10

US-10-037-417-46 (1-357) x BF561257 (1-296)
 QY 55 ProGlyThrTrpProTglnValSerLeuHisGlyGlyHisileCysGlyGly 74
 Db 270 CCTGGCACTTGGCGCTGGAGCTGCATCAGCGTGAGGCCACATCTGGGGGGC 211
 QY 75 SerLeuileAlaProSerTrpValSerLeuHisCysPhe 89
 Db 210 TCCCTCATCGCCCTTCTCGGCTCTCTCCGCTGCTACTGTTTC 166

RESULT 5
 AA996811
 LOCUS
 DEFINITION
 Rattus norvegicus (Norway rat)
 SOURCE
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus
 1 (bases 1 to 322)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 On Jun 5, 1998 this sequence version replaced gi:3187366.
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250

Alignment Scores: 5.11e-21 Length: 322
 Pred. No.: 35.00 Matches: 35
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 9.80% Gaps: 0
 DB: 9

US-10-037-417-46 (1-357) x AA996811 (1-322)
 QY 55 ProGlyThrTrpProTglnValSerLeuHisGlyGlyHisileCysGlyGly 74
 Db 50 CCTGGCACTTGGCGCTGGAGCTGCATCAGCGTGAGGCCACATCTGGGGGGC 109
 QY 75 SerLeuileAlaProSerTrpValSerLeuHisCysPhe 89
 Db 110 TCCCTCATCGCCCTTCTCGGCTCTCTCCGCTGCTACTGTTTC 154

RESULT 6
 BY236668
 LOCUS
 DEFINITION
 Rattus norvegicus (Norway rat)
 SOURCE
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus
 1 (bases 1 to 322)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 On Jun 5, 1998 this sequence version replaced gi:3187366.
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250

12466851
 PUBMED
 COMMENT
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp
 URL: http://genome.gsc.riken.go.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Kirk W. Beisel (Boys Town National
 Research Hospital 555 North 30th Street Omaha, NE 68131 USA) whose
 assistance we gratefully acknowledge. Please visit our web site
 (http://genome.gsc.riken.go.jp) for further details.
 Location/Qualifiers
 source 1. 365
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="F930045I01"
 /tissue_type="inner ear"
 /dev_stage="adult"
 /clone_lib="RIKEN full-length enriched, adult inner ear"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.88e-10 Length: 365
 Score: 23.00 Matches: 23
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.44% Indels: 0
 DB: 13 Gaps: 0
 US-10-037-417-46 (1-357) x BY235958 (1-365)
 QY 67 GlyGlyGlyHisIleCysGlyGlySerIleuIleAlaProSerTrpValIeuSerIala 86
 Db 265 GTGGGGGCCCATCTGCGGGGGCTCCCTCATCGCCCTTCCTCGGTCTCCGCTGCT 324
 QY 87 HisCysPhe 89
 Db 325 CACTGTTC 333
 RESULT 8
 BY236911
 LOCUS
 DEFINITION
 BY236911 RIKEN full-length enriched, adult inner ear Mus musculus
 CDNA clone F930050D15 5', mRNA sequence.
 ACCESSION
 BY236911
 VERSION
 BY236911.1 GI:26418099

KEYWORDS

SOURCE

ORGANISM

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 369)

REFERENCE

AUTHORS

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 369)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, T.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusica, V.,
 Chothia, C., Corbani, L. E., Cousins, S., Dalia, E., Dragani, T. A.,
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
 Gariboldi, M., Glasi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
 Kurochi, H., Kawasawa, Y., Kiedziera, R. M., King, B. L., Konagaya, A.,
 Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W. J., Pertea, G., Resole, G.,
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp

URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,

Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,

Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,

Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,

Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Tissues were provided by Kirk W. Beisel (Boys Town National

Research Hospital 555 North 30th Street Omaha, NE 68131 USA) whose

assistance we gratefully acknowledge. Please visit our web site

(http://genome.gsc.riken.go.jp) for further details.

Thu Feb 26 13:50:26 2004

us-10-037-417-46.olip2n.rst

```

FEATURES
  source
    Location/Qualifiers
      1..369
        /organism="Mus musculus"
        /mol_type="mRNA"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="F930050D15"
        /tissue_type="inner ear"
        /dev_stage="adult"
        /clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN
Alignment Scores:
  Pred. No.: 2,92e-10 Length: 369
  Score: 23.00 Matches: 23
  Percent Similarity: 100.00% Conservatives: 0
  Best Local Similarity: 100.00% Mismatches: 0
  Query Match: 6.44% Indels: 0
  DB: 13 Gaps: 0

US-10-037-417-46 (1-357) x BY236911 (1-369)
Qy 67 GlyGlyHisIleCyseGlySerLeuIleAlaProSerTrpValLeuSerAlaAla 86
Db 271 GGTGGGGCCACATCTGGGGGGCTCCCTCATCGCCCTTCTGGGTCTCTCGCTGCT 330
Qy 87 HisCysPhe 89
Db 331 CACTGTTTC 339

RESULT 9
AZ252377
LOCUS
  RPCI-23-466N15.TVB RPCI-23 Mus musculus genomic clone
  RPCI-23-466N15, genomic survey sequence.
ACCESSION
  AZ252377
VERSION
  AZ252377.1 GI:8565580
KEYWORDS
  GSS.
SOURCE
  Mus musculus (house mouse)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE
    1 (bases 1 to 430)
    Zhao, S., Nierman, W., Feidblyum, T., Malek, J., Shatsman, S.,
    Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de
    Jong, P. and Fraser, C.M.
    Mouse BAC End Sequences from Library RPCI-23
    Unpublished (1999)
    Other GSSs: RPCI-23-466N15.TJ RPCI-23-466N15.TV RPCI-23-466N15.TVB
    Contact: Shaying Zhao
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850, USA
    Tel: 301 838 0200
    Fax: 301 838 0208
    Email: szhao@tigr.org
    Clones are derived from the mouse BAC library RPCI-23. For BAC
    library availability, please contact Pieter de Jong
    (pieter@dejong.med.buffalo.edu). Clones may be purchased from
    BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
    or from Resea ch Genetics (info@resgen.com). BAC end page:
    http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
    Plate: 466 row: N column: 15
    Seq primer: SP6
    Class: BAC ends.
    Location/Qualifiers
      1..430
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="RPCI-23-466N15"
        /sex="Female"

FEATURES
  source
    /lab_host="DH10B"
    /clone_lib="RPCI-23"
    /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
    EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
    brain genomic DNA was isolated and partially digested
    with a combination of EcoRI and EcoRI Methyase. Size
    selected DNA was cloned into the pBACe3.6 vector at the
    EcoRI sites. The ligation products were transformed into
    DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN
Alignment Scores:
  Pred. No.: 3,46e-10 Length: 430
  Score: 23.00 Matches: 23
  Percent Similarity: 100.00% Conservatives: 0
  Best Local Similarity: 100.00% Mismatches: 0
  Query Match: 6.44% Indels: 0
  DB: 28 Gaps: 0

US-10-037-417-46 (1-357) x AZ252377 (1-430)
Qy 258 AlaGlyIleThrSerPheGlyPheGlyCysGlyArgAsnArgProGlyValpHeThr 277
Db 93 GCCGGAATCACCACTTTGGCTTTGGCTTGTGGACGAGGACCGCCCTGGGGTCTTCACT 152
Qy 278 AlaValAla 280
Db 153 GCAGTGGCT 161

RESULT 10
BB850607
LOCUS
  BB850607 RIKEN full-length enriched, adult inner ear Mus musculus
  CDNA clone F930108E12 5', mRNA sequence.
ACCESSION
  BB850607
VERSION
  BB850607.1 GI:17092061
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE
    1 (bases 1 to 433)
    Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
    Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K.,
    Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
    Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,
    Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,
    Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
    Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
    Tanaka, T., Tonaru, A., Toya, T., Watahiki, A., Yasunishi, A.,
    Muramatsu, M. and Hayashizaki, Y.
    RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
    2001)
    Unpublished (2001)
    Contact: Yoshihide Hayashizaki
    Laboratory for Genome Exploration Research Group, RIKEN Genomic
    Sciences Center (GSC), Yokohama Institute
    The Institute of Physical and Chemical Research (RIKEN)
    1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
    Tel: 81-45-503-9222
    Fax: 81-45-503-9216
    Email: genome-res@sc.riken.go.jp
    URL: http://genome.gsc.riken.go.jp/
    Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
    Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
    Normalization and subtraction of cap-trapper-selected cDNAs to
    prepare full-length cDNA libraries for rapid discovery of new
    genes. Genome Res. 10 (10), 1617-1630 (2000)
    wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
    Watahiki, M., Itoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
    Matsuura, S., Kawai, Y., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.
    and Hayashizaki, Y.
    RIKEN integrated sequence analysis (RISA) system--384-format

```

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11) 1757-1771 (2000)
 Komo,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
 source
 Location/Qualifiers
 1. .433
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="F930108E12"
 /tissue_type="inner ear"
 /dev_stage="adult"
 /clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN
 Alignment Scores:
 Pred. No.: 3.49e-10 Length: 433
 Score: 23.00 Matches: 23
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.44% Indels: 0
 DB: 10 Gaps: 0

US-10-037-417-46 (1-357) x BB850607 (1-433)

Qy 67 GlyGlyGlyHisIleCysGlySerLeuIleAlaProSerTrpValLeuSerAlaAla 86
 Db 274 GGTGGGGGGGCACATCTGCGGGGGCTCCCTCATGCGCCCTCTCGGGTCTCTCGGTGCT 333

Qy 87 HisCysPhe 89
 Db 334 CACTGTTC 342

RESULT 11
 AZ875648/c 454 bp DNA linear GSS 21-FEB-2001
 LOCUS
 DEFINITION
 2M0190113F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0190113 F, genomic survey sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)

REFERENCE
 AUTHORS
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Isalam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,R., Stokes,R., Tinge,A., von Niederhausen,A. and Wright,D.,Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL
 COMMENT
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 309, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0190 row: 1 column: 13
 Seq primer: CTTTGAACGACGGCCAGT
 Claes: plasmid ends
 High quality sequence stop: 454.

FEATURES
 source
 Location/Qualifiers
 1. .454

/organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0190113"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Alignment Scores:
 Pred. No.: 3.68e-10 Length: 454
 Score: 23.00 Matches: 23
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.44% Indels: 0
 DB: 28 Gaps: 0

US-10-037-417-46 (1-357) x AZ875648 (1-454)

Qy 67 GlyGlyGlyHisIleCysGlySerLeuIleAlaProSerTrpValLeuSerAlaAla 86
 Db 271 GGTGGGGGGGCACATCTGCGGGGGCTCCCTCATGCGCCCTCTCGGTGCTCTCGGTGCT 212
 Qy 87 HisCysPhe 89
 Db 211 CACTGTTC 203

RESULT 12

BB850564 433 bp mRNA linear EST 26-NOV-2001
 LOCUS
 DEFINITION
 BB850564 RIKEN full-length enriched, adult inner ear Mus musculus cDNA clone F930108A18 5', mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)

REFERENCE
 AUTHORS
 Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Inotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL
COMMENT

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Explore Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.,
A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001). Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
source

```

location/qualifiers
1. 433
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F930108A18"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-l

```

```
ORIGIN
/clone_lab="RIKEN full-length enriched, adult inner ear"
```

Alignment Scores:	
Pred. No.:	1.64e-07
Score:	20.00
Length:	433
Matches:	20
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0
Query Match:	5.60%
Best Local Similarity:	100.00%
Percent Similarity:	100.00%

US-10-037-417-46 (1-357) x BB850564 (1-433)

67 GlyGlyGlyHisIleCysGlyGlySerIleuIleAlaProSerTrpValLeuSerAlaAla 86
274 GGTGGGGCCACATCTGGGGGGTCCCTCATGCCCTTCTTGGGGTCTTCCGTGCT 333

RESULT 13

BU164841/C
 LOCUS
 DEFINITION
 BU164841
 AGENCEOUT_7906304 NTH_MGC_67 Homo sapiens
 5', mRNA sequence.
 BU164841
 BU164841.1 GI:22678793
 EST.
 Homo sapiens (human)
 Homo sapiens
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 875)
 NIH-MGC <http://mgc.ncbi.nlm.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 UNPUBLISHED (1999)
 CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13459 row: d column: 24
High quality sequence stop: 645.

FEATURES	source
----------	--------

1. ...875
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6149519"
/tissue_type="retinoblastoma"
/lab_host="PH10B (phage-resistant)"
/clone_lib="NH16667"
/note="Organ: eye; Vector: pCMV-SPO1"
Size: 3.6 kb; Cloned unidirectional.
Average insert size 1.75 kb. Library
Technologies."

ORIGIN

Alignment Scores.

Pred. No.:	0.00017	Length:	875
Score:	17.00	Matches:	17
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.76%	Indels:	0
DB:	13	Gaps:	0

US-10-037-417-46 (1-357) x BUL64841 (1-875)

341 LeuValLeuProTIPLeuSerProHisSerLeuLeuGlyLeuTrpGlyPhe 357
QY |||||
668 CTTCCTCTCCCTGGCTCTCTCACACTCTCTCTGGGCTCTGGGGTTT 618
Db |||||

RESULT 14

CD251990/c	CD251990	907 bp	mrna	linear	EST 22-MAY-2003
LOCUS	AGENCOURT 14211954	NIH_MGC_180	Homo sapiens	cdna clone	
DEFINITION	IMAGE:30383555 5',	mrna	sequence.		
ACCESSION	CD251990				
VERSION	CD251990.1	GI:31012456			
KEYWORDS	EST.				
ORGANISM	Homo sapiens	(human)			
	Eukaryota; Metazoa;	Chordata; Craniata;	Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria;	Primates; Catarrhini;	Hominidae; Homo.		
AUTHORS	1. (bases 1 to 907)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Daniela S. Gerhard.	Ph.D.			

National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20899
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Dr. Michael Br
 cDNA Library Preparation: Invitro
 cDNA Library Arrayed by: The I.M.
 DNA sequencing by: Agencourt Bios
 Clone distribution: MGC clone dis
 found through the I.M.A.G.E. Conso
 http://image.llnl.gov
 Plate: NDAM447 row: j column: 12
 High quality sequence stop: 548.
 Location/Qualifiers
 1. .907
 /organism="Homo sapiens"
 /mol_type="mRNA"
 FEATURES
 source

/db_xref=taxon:9606"
/clone="IMAGE:30383555"
/lab_host="DH10B-fcn A (T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_180"
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
Site 2: EcoRV (destroyed); Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC
Library."

ORIGIN

Alignment Scores:
Pred. No.: 0.000177 Length: 907
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.76% Indels: 0
DB: 14 Gaps: 0

US-10-037-417-46 (1-357) x CD2511990 (1-907)

QY 341 LeuValLeuProTriPLeuSerProHisSerLeuLeuGlyLeuTriPglyPhe 357
Db 520 CTGTGCTCCCGGCTCTCTCCACACTCTCTCTGGGCTCTGGGGGTTT 470

RESULT 15
BU860221/c
LOCUS
DEFINITION AGENCOURT.10435504 NIH_MGC_107 Homo sapiens cDNA clone
IMAGE:6651279 5', mRNA sequence.
ACCESSION BU860221
VERSION BU860221.1 GI:24045213
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 929)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM2897 row: n column: 15
High quality sequence stop: 423.
Location/Qualifiers
1. .929
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6651279"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_107"
/note="Organ: breast; Vector: pOTB7; Site 1: EcoRI;
Site 2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRV/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 0.000182 Length: 929
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.76% Indels: 0
DB: 13 Gaps: 0
US-10-037-417-46 (1-357) x BU860221 (1-929)
QY 341 LeuValLeuProTriPLeuSerProHisSerLeuLeuGlyLeuTriPglyPhe 357
Db 520 CTGTGCTCCCGGCTCTCTCCACACTCTCTCTGGGCTCTGGGGGTTT 2

Search completed: February 26, 2004, 00:06:29
Job time : 2502 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 25, 2004, 18:59:12 ; Search time 97 Seconds

(without alignments)

2042.448 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 357

Sequence: 1 MAQKGVLPGLGAVANSDS.....TKSLVLPWLSPLHLLGLMGF 357

Scoring table:

	OLIGO	Xgapop 60.0	Xgapext 60.0
		Ygapop 60.0	Ygapext 60.0
		Fgapop 6.0	Fgapext 7.0
		Delop 6.0	Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1360453

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10037417/runat.24022004.101109.18111/app.query.fasta_1.519
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=olip2n.rni -MINWATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MAIRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10037417 @CGN 1.1 103 @runat.24022004.101109.18111 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPHLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:**

	1:	/cgn2_6/ptodata/2/ina/5A.COMB.seq:**
	2:	/cgn2_6/ptodata/2/ina/5B.COMB.seq:**
	3:	/cgn2_6/ptodata/2/ina/6A.COMB.seq:**
	4:	/cgn2_6/ptodata/2/ina/6B.COMB.seq:**
	5:	/cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:**
	6:	/cgn2_6/ptodata/2/ina/backfiles1.seq:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	13	3.6	506	4	US-09-280-116-72
2	13	3.6	792	4	US-09-244-111-3
3	13	3.6	944	3	US-09-070-526-1
4	13	3.6	994	3	US-09-008-271A-19
5	13	3.6	1049	4	US-09-386-642-9
C 6	13	3.6	1343	4	US-09-618-259-72
7	13	3.6	1360	4	US-09-618-259-6
8	11	3.1	262	3	US-08-944-483-5
C 9	11	3.1	340	4	US-09-280-116-28
C 10	11	3.1	340	4	US-09-280-116-29
11	11	3.1	436	3	US-08-906-769-122
12	11	3.1	436	3	US-08-906-616-122

13	11	3.1	436	3	US-08-639-075A-122	Sequence 122, App
14	11	3.1	436	3	US-09-012-431-122	Sequence 122, App
15	11	3.1	436	3	US-09-032-215-19	Sequence 19, Appl
16	11	3.1	436	3	US-09-012-692-122	Sequence 122, App
17	11	3.1	436	3	US-08-906-613-122	Sequence 122, App
18	11	3.1	453	6	5200340-3	Patent No. 5200340
19	11	3.1	662	4	US-09-702-705-109	Sequence 109, App
20	11	3.1	662	4	US-09-736-457-109	Sequence 109, App
21	11	3.1	662	4	US-09-614-124B-109	Sequence 109, App
22	11	3.1	662	4	US-09-671-325-109	Sequence 109, App
23	11	3.1	662	4	US-09-589-184-109	Sequence 109, App
24	11	3.1	690	1	US-08-379-621-1	Sequence 1, Appli
25	11	3.1	690	2	US-08-889-078-1	Sequence 1, Appli
26	11	3.1	732	1	US-08-361-395-2	Sequence 2, Appli
27	11	3.1	735	3	US-09-079-970A-1	Sequence 1, Appli
28	11	3.1	771	3	US-09-079-970A-4	Sequence 4, Appli
29	11	3.1	833	2	US-08-790-137-2	Sequence 2, Appli
30	11	3.1	840	4	US-09-244-111-5	Sequence 5, Appli
31	11	3.1	933	4	US-09-023-942A-29	Sequence 29, Appli
32	11	3.1	1031	2	US-08-978-404B-1	Sequence 1, Appli
33	11	3.1	1052	4	US-09-386-642-10	Sequence 10, Appli
34	11	3.1	1065	1	US-08-427-640-1	Sequence 1, Appli
35	11	3.1	1065	2	US-08-427-640-5	Sequence 5, Appli
36	11	3.1	1065	2	US-08-811-949-60	Sequence 60, Appli
37	11	3.1	1068	1	US-08-137-116-2	Sequence 2, Appli
38	11	3.1	1068	1	US-08-427-640-3	Sequence 3, Appli
39	11	3.1	1068	1	US-08-427-640-7	Sequence 7, Appli
40	11	3.1	1068	2	US-08-811-949-44	Sequence 44, Appli
41	11	3.1	1068	2	US-08-811-949-46	Sequence 46, Appli
42	11	3.1	1068	2	US-08-811-949-52	Sequence 52, Appli
43	11	3.1	1068	2	US-08-811-949-58	Sequence 58, Appli
44	11	3.1	1068	6	5223256-3	Patent No. 5223256
45	11	3.1	1081	2	US-09-016-366A-22	Sequence 22, Appli

ALIGNMENTS

RESULT 1

US-09-280-116-72/c

; Sequence 72, Application US/09280116A

; Patent No. 6331427

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs

; FILE REFERENCE: 5800-24, 035800/176965

; CURRENT APPLICATION NUMBER: US/09/280,116A

; CURRENT FILING DATE: 1999-03-26

; NUMBER OF SEQ ID NOS: 268

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 72

; LENGTH: 506

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: trypsin-like serine proteases

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)-(506)

; OTHER INFORMATION: n = a, t, c or g

US-09-280-116-72

Alignment Scores:			
Pred. No.:	0.000809	Length:	506
Score:	13.00	Matches:	13
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64%	Indels:	0
DB:	4	Gaps:	0

US-10-037-417-46 (1-357) x US-09-280-116-72 (1-506)

Qy 239 CysGlnGlyAspSerGlyProLeuValCysGlu 251

|||||

Db 329 TCCAGGGTGATTCGGAGGCCGCTGGTGTGAGGAA 291
RESULT 2
US-09-244-111-3
; Sequence 3, Application US/09244111
; Patent No. 6566498
; GENERAL INFORMATION:
; APPLICANT: Ni, et al.
; TITLE OF INVENTION: Human Serine Protease and Serpin Polypeptides
; FILE REFERENCE: PF391
; CURRENT APPLICATION NUMBER: US/09/244,111
; EARLIER FILING DATE: 1999-02-04
; EARLIER APPLICATION NUMBER: 60/073,961
; EARLIER FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 3
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (85)..(708)
US-09-244-111-3
Alignment Scores:
Pred. No.: 0.00125 Length: 792
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 4 Gaps: 0
US-10-037-417-46 (1-357) x US-09-244-111-3 (1-792)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeuValCys 249
Db 541 GACACGTGCCAGGCGATTCTGGAGGCCCTCGTGTGT 579
RESULT 3
US-09-070-526-1
; Sequence 1, Application US/09070526
; Patent No. 6100059
; GENERAL INFORMATION:
; APPLICANT: SOUTHAN, CHRISTOPHER
; APPLICANT: CLINKENBEARD, HELEN
; APPLICANT: BURGESS, NICOLA
; TITLE OF INVENTION: NO. 6100059e1 Compounds
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,526
; FILING DATE: 30-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9711952.3
; FILING DATE: 9-JUN-1997
; APPLICATION NUMBER: EP 97309646.4
; FILING DATE: 1-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031

; REFERENCE/DOCKET NUMBER: GH-30353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 944 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-070-526-1
Alignment Scores:
Pred. No.: 0.00149 Length: 944
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 3 Gaps: 0
US-10-037-417-46 (1-357) x US-09-070-526-1 (1-944)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeuValCys 249
Db 720 GACACGTGCCAGGCGATTCTGGAGGCCCTCGTGTGT 758
RESULT 4
US-09-008-271A-19
; Sequence 19, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 994 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```
/
/ IMMEDIATE SOURCE:
/ LIBRARY: COLNROT27
/ CLONE: 1798496
/ SEQUENCE DESCRIPTION: SEQ ID NO: 19 :
US-09-008-271A-19

Alignment Scores:
Pred. No.: 0.00156 Length: 994
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 3 Gaps: 0

US-10-037-417-46 (1-357) x US-09-008-271A-19 (1-994)

Qy 237 AspThrCysGlnGlyAspSerGlyClyProLeuValCys 249
Db 770 GACACGTGCCAGGCGGATCTGGAGGCCCTCGGTGTGT 808

RESULT 5
US-09-386-642-9
; Sequence 9, Application US/09386642
; Patent No. 6420157
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jensen
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Zymogen Activation System
; FILE REFERENCE: ORI-1028
; CURRENT APPLICATION NUMBER: US/09/386,642
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1049
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
; OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-9

Alignment Scores:
Pred. No.: 0.00165 Length: 1049
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 4 Gaps: 0

US-10-037-417-46 (1-357) x US-09-386-642-9 (1-1049)

Qy 237 AspThrCysGlnGlyAspSerGlyClyProLeuValCys 249
Db 688 GACACGTGCCAGGCGGATCTGGAGGCCCTCGGTGTGT 726

RESULT 6
US-09-618-259-72/c
; Sequence 72, Application US/09618259
; Patent No. 6642013
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: No. 6642013el Extracellular Serine Protease
; FILE REFERENCE: D6020CIP2
; CURRENT APPLICATION NUMBER: US/09/618,259
; CURRENT FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: US 09/127,444
; PRIOR FILING DATE: 1998-08-21
; NUMBER OF SEQ ID NOS: 72
; SEQ ID NO 72
; LENGTH: 1343
```

```
/
/ TYPE: DNA
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Anti-sense sequence of TADG-14
US-09-618-259-72

Alignment Scores:
Pred. No.: 0.0021 Length: 1343
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 4 Gaps: 0

US-10-037-417-46 (1-357) x US-09-618-259-72 (1-1343)

Qy 237 AspThrCysGlnGlyAspSerGlyClyProLeuValCys 249
Db 213 GACACGTGCCAGGCGGATCTGGAGGCCCTCGGTGTGT 175

RESULT 7
US-09-618-259-6
; Sequence 6, Application US/09618259
; Patent No. 6642013
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: No. 6642013el Extracellular Serine Protease
; FILE REFERENCE: D6020CIP2
; CURRENT APPLICATION NUMBER: US/09/618,259
; CURRENT FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: US 09/127,444
; PRIOR FILING DATE: 1998-08-21
; NUMBER OF SEQ ID NOS: 72
; SEQ ID NO 6
; LENGTH: 1360
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence encoding Tumor Antigen Derived
; OTHER INFORMATION: Gene-14 (TADG-14) protein; nt 1344-1360 NCBI accession
; OTHER INFORMATION: #AA343629
US-09-618-259-6

Alignment Scores:
Pred. No.: 0.00213 Length: 1360
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 4 Gaps: 0

US-10-037-417-46 (1-357) x US-09-618-259-6 (1-1360)

Qy 237 AspThrCysGlnGlyAspSerGlyClyProLeuValCys 249
Db 1131 GACACGTGCCAGGCGGATCTGGAGGCCCTCGGTGTGT 1169

RESULT 8
US-08-944-483-5
; Sequence 5, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; OTHER INFORMATION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
```


;/ CURRENT APPLICATION DATA: US/08/906,769
;/ APPLICATION NUMBER: US/08/906,769
;/ FILING DATE:

CLASSIFICATION:

;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/639,075
;/ FILING DATE: 24-APR-1996

ATTORNEY/AGENT INFORMATION:

;/ NAME: Connell, Gary J.
;/ REGISTRATION NUMBER: 32,020
;/ REFERENCE/DOCKET NUMBER: 2618-25-C2

TELECOMMUNICATION INFORMATION:

;/ TELEPHONE: (303) 863-9700

;/ TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 122:

SEQUENCE CHARACTERISTICS:

;/ LENGTH: 436 base pairs

;/ TYPE: nucleic acid

;/ STRANDEDNESS: single

;/ TOPOLOGY: linear

;/ MOLECULE TYPE: cDNA

FEATURE:

;/ NAME/KEY: CDS

;/ LOCATION: 2..436

;/ OTHER INFORMATION: /note= "At pos. bp 301, change A to

;/ OTHER INFORMATION: W; at pos. bp 342, change C to Y; at pos. bp 397, change C to
;/ OTHER INFORMATION: at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,
;/ OTHER INFORMATION: substitute Xaa."

US-08-906-769-122

Alignment Scores:

Pred. No.:	0.0669	Length:	436
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.08%	Indels:	0
DB:	3	Gaps:	0

US-10-037-417-46 (1-357) x US-08-906-769-122 (1-436)

Oy 237 AspThrCysGlnGlyAspSerGlyClyProLeu 247

Db 257 GACACTTCCAGGAGATTTCAGGTGGCCATTG 289

RESULT 12

US-08-906-616-122

;/ Sequence 122, Application US/08906616

;/ Patent No. 6121035

GENERAL INFORMATION:

;/ APPLICANT: Grieve, Robert B.

;/ APPLICANT: Rushlow, Keith E.

;/ APPLICANT: Wu Hunter, Shirley

;/ APPLICANT: Frank, Glenn R.

;/ APPLICANT: Stiegler, Gary

;/ APPLICANT: Gaines, Patrick J.

;/ APPLICANT: Silver, Gary

;/ TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF

;/ NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

;/ ADDRESSEE: Sheridan Ross P.C.

;/ STREET: 1700 Lincoln Street, Suite 3500

;/ CITY: Denver

;/ STATE: Colorado

;/ COUNTRY: USA

;/ ZIP: 80203

COMPUTER READABLE FORM:

;/ MEDIUM TYPE: Floppy disk

;/ COMPUTER: IBM PC compatible

;/ OPERATING SYSTEM: PC-DOS/MS-DOS

;/ SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

;/ APPLICATION NUMBER: US/08/906,616

;/ FILING DATE: 05-AUG-1997

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

;/ NAME: Connell, Gary J.

;/ REGISTRATION NUMBER: 32,020

;/ REFERENCE/DOCKET NUMBER: 2618-25-C2-3

TELECOMMUNICATION INFORMATION:

;/ TELEPHONE: (303) 863-9700

;/ TELEFAX: (303) 863-0223

;/ INFORMATION FOR SEQ ID NO: 122:

SEQUENCE CHARACTERISTICS:

;/ LENGTH: 436 base pairs

;/ TYPE: nucleic acid

;/ STRANDEDNESS: single

;/ TOPOLOGY: linear

;/ MOLECULE TYPE: cDNA

FEATURE:

;/ NAME/KEY: CDS

;/ LOCATION: 2..436

;/ OTHER INFORMATION: /note= "At pos. bp 301, change A to

;/ OTHER INFORMATION: W; at pos. bp 342, change C to Y; at pos. bp 397, change C to
;/ OTHER INFORMATION: at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,
;/ OTHER INFORMATION: substitute Xaa."

US-08-906-616-122

Alignment Scores:

Pred. No.:	0.0669	Length:	436
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.08%	Indels:	0
DB:	3	Gaps:	0

US-10-037-417-46 (1-357) x US-08-906-616-122 (1-436)

Oy 237 AspThrCysGlnGlyAspSerGlyClyProLeu 247

Db 257 GACACTTCCAGGAGATTTCAGGTGGCCATTG 289

RESULT 13

US-08-639-075A-122

;/ Sequence 122, Application US/08639075A

;/ Patent No. 6150125

GENERAL INFORMATION:

;/ APPLICANT: Grieve, Robert B.

;/ APPLICANT: Rushlow, Keith E.

;/ APPLICANT: Wu Hunter, Shirley

;/ APPLICANT: Frank, Glenn R.

;/ APPLICANT: Stiegler, Gary

;/ APPLICANT: Gaines, Patrick J.

;/ APPLICANT: Silver, Gary

;/ TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID

;/ NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

;/ ADDRESSEE: Sheridan Ross & McIntosh

;/ STREET: 1700 Lincoln Street, Suite 3500

;/ CITY: Denver

;/ STATE: Colorado

;/ COUNTRY: USA

;/ ZIP: 80203

COMPUTER READABLE FORM:

;/ MEDIUM TYPE: Floppy disk

;/ COMPUTER: IBM PC compatible

;/ OPERATING SYSTEM: PC-DOS/MS-DOS

;/ SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

;/ APPLICATION NUMBER: US/08/639,075A

;/ FILING DATE: 24-APR-1996

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

;/ NAME: Connell, Gary J.

;/ REGISTRATION NUMBER: 32,020

;/ REFERENCE/DOCKET NUMBER: 2618-25-C2


```
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (303) 863-9700
/ TELEFAX: (303) 863-0223
/ INFORMATION FOR SEQ ID NO: 122:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 436 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 2..436
/ OTHER INFORMATION: /note= "at pos. bp 301, change A to
/ W; at pos. bp 342, change C to Y; at pos. bp 397, change C to Y;
/ at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,
/ substitute Xaa."
/ SEQUENCE DESCRIPTION: SEQ ID NO: 122:
US-09-012-431-122

Alignment Scores:
Pred. No.: 0.0669 Length: 436
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.08% Indels: 0
DB: 3 Gaps: 0

US-10-037-417-46 (1-357) x US-09-012-431-122 (1-436)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeu 247
DB 257 GACACTTGCAGGAGATTTCAGGTGGGCCATTG 289

RESULT 15
US-09-032-215-19
; Sequence 19, Application US/09032215
; Patent No. 6204010
; GENERAL INFORMATION:
; APPLICANT: Stiegler, Gary L.
; APPLICANT: Gaines, Patrick J.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: .50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C. Suite 3500
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII DOS TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,215
; FILING DATE: 27-FEB-1998
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (303) 863-9700
/ TELEFAX: (303) 863-0223
/ INFORMATION FOR SEQ ID NO: 122:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 436 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 2..436
/ OTHER INFORMATION: /note= "at pos. bp 301, change A to
/ W; at pos. bp 342, change C to Y; at pos. bp 397, change C to
/ Y; at pos. bp 431, change G to S. At pos. aa 100, 114 and 144,
/ substitute Xaa."
/ SEQUENCE DESCRIPTION: substitute Xaa."
US-08-639-075A-122

Alignment Scores:
Pred. No.: 0.0669 Length: 436
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.08% Indels: 0
DB: 3 Gaps: 0

US-10-037-417-46 (1-357) x US-08-639-075A-122 (1-436)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeu 247
DB 257 GACACTTGCAGGAGATTTCAGGTGGGCCATTG 289

RESULT 14
US-09-012-431-122
; Sequence 122, Application US/09012431
; Patent No. 6180383
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/012,431
; FILING DATE: 23-Jan-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..436
; FEATURE:
; NAME/KEY: W = A
; LOCATION: 301
; FEATURE:
; NAME/KEY: Y = C
; LOCATION: 342, 397
; FEATURE:
; NAME/KEY: S = C
; LOCATION: 431
; FEATURE:
; NAME/KEY: Xaa = Unknown
; LOCATION: 100, 114, 144
US-09-032-215-19

```

```

Alignment Scores:
Pred. No.: 0.0669 Length: 436
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.08% Indels: 0
DB: 3 Gaps: 0

```

US-10-037-417-46 (1-357) x US-09-032-215-19 (1-436)

```

Qy 237 AspThrCysGlnGlyAspSerGlyGlyProLeu 247
Db 257 GACACTGCCAGGGAGATTTCAGTGGGCATTG 289

```

Search completed: February 26, 2004, 00:08:19
Job time : 101 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2004, 23:28:21 ; Search time 399 Seconds

(without alignments)
3133.050 Million cell updates/sec

Title: US-10-037-417-46

Perfect score: 357

Sequence: 1 MAQGVLPQQLGAVANSDS.....TKSLVLPWLSPHSLGLWGF 357

Scoring table:

OLIGO	
Xgapop 60.0 , Xgapext 60.0	
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2308684 seqs, 1750822206 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4610896

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+, p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10037417/runat_24022004_101110_18195/app_query.fasta_1.519
-DB=Published Applications NA -SUFFIX=olip2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1
-TRANS=human40.cdi -UNIT=bits -START=1 -END=1 -MATRIX=oligo
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10037417 @CGN 1 1 107 @runat_24022004_101110_18195
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
17:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
18:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

1	272	76.2	2457	9	US-09-888-615-52	Sequence 52, Appl
2	45	12.6	768	14	US-10-221-097-10	Sequence 10, Appl
3	16	4.5	315	14	US-10-029-386-21306	Sequence 21306, A
4	16	4.5	543	14	US-10-029-386-7606	Sequence 7606, Ap
5	16	4.5	1161	12	US-10-042-865-31	Sequence 31, Appl
6	16	4.5	1726	12	US-10-042-865-29	Sequence 29, Appl
7	16	4.5	1733	14	US-10-176-847-85	Sequence 85, Appl
8	16	4.5	1834	9	US-09-948-094-1	Sequence 1, Appli
9	16	4.5	1834	9	US-09-880-107-2214	Sequence 2214, Ap
10	16	4.5	1834	14	US-09-967-768A-141	Sequence 141, App
11	16	4.5	1834	14	US-10-097-340-261	Sequence 261, App
12	16	4.5	3382	14	US-10-101-510-447	Sequence 447, App
13	13	3.6	792	14	US-10-325-745-3	Sequence 3, Appli
14	13	3.6	988	14	US-10-101-510-621	Sequence 621, App
15	13	3.6	994	14	US-10-101-510-612	Sequence 612, App
16	13	3.6	994	14	US-10-180-719-19	Sequence 19, Appl
17	13	3.6	999	9	US-09-978-295A-394	Sequence 394, App
18	13	3.6	999	9	US-09-978-697-394	Sequence 394, App
19	13	3.6	999	9	US-09-978-192A-394	Sequence 394, App
20	13	3.6	999	9	US-09-999-832A-394	Sequence 394, App
21	13	3.6	999	10	US-09-978-189-394	Sequence 394, App
22	13	3.6	999	10	US-09-978-608A-394	Sequence 394, App
23	13	3.6	999	10	US-09-978-585A-394	Sequence 394, App
24	13	3.6	999	10	US-09-978-191A-394	Sequence 394, App
25	13	3.6	999	10	US-09-978-403A-394	Sequence 394, App
26	13	3.6	999	10	US-09-978-564A-394	Sequence 394, App
27	13	3.6	999	10	US-09-999-833A-394	Sequence 394, App
28	13	3.6	999	10	US-09-981-915A-394	Sequence 394, App
29	13	3.6	999	10	US-09-978-824-394	Sequence 394, App
30	13	3.6	999	10	US-09-918-585A-394	Sequence 394, App
31	13	3.6	999	10	US-09-978-423A-394	Sequence 394, App
32	13	3.6	999	10	US-09-978-193A-394	Sequence 394, App
33	13	3.6	999	10	US-09-999-830A-394	Sequence 394, App
34	13	3.6	999	10	US-09-978-757A-394	Sequence 394, App
35	13	3.6	999	10	US-09-978-187B-394	Sequence 394, App
36	13	3.6	999	10	US-09-978-643A-394	Sequence 394, App
37	13	3.6	999	10	US-09-978-375A-394	Sequence 394, App
38	13	3.6	999	10	US-09-978-298A-394	Sequence 394, App
39	13	3.6	999	10	US-09-978-188A-394	Sequence 394, App
40	13	3.6	999	10	US-09-978-681A-394	Sequence 394, App
41	13	3.6	999	10	US-09-978-194A-394	Sequence 394, App
42	13	3.6	999	10	US-09-999-829A-394	Sequence 394, App
43	13	3.6	999	10	US-09-978-299A-394	Sequence 394, App
44	13	3.6	999	10	US-09-978-544A-394	Sequence 394, App
45	13	3.6	999	10	US-09-978-665A-394	Sequence 394, App

ALIGNMENTS

RESULT 1
US-09-888-615-52
; Sequence 52, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:

; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHITE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 52
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-888-615-52

Alignment Scores:
Pred. No.: 1.57e-260 Length: 2457
Score: 272.00 Matches: 272
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.13% Indels: 0
DB: 9 Gaps: 0

US-10-037-417-46 (1-357) x US-09-888-615-52 (1-2457)

QY 38 CysGlyArgProGluProSerAlaAlaValGlyGlySerAsnAlaGlnProGlyThr 57
DB 112 TGCGGGCGCCCTGAGCCCTCGCCCGCATCGTGGGGGCTCAACGCGCAGCGGGCACC 171
QY 58 TrpProTrpGlnValSerLeuHisHisGlyGlyHisLeCysGlyGlySerLeu 77
DB 172 TGGCCCTGGCAAGTGAAGCTGCACCATGAGGTGGCCCATCTGGGGGGGCTCCCTCATC 231
QY 78 AlaProSerTrpValLeuSerAlaAlaHisCysPheMetThrAsnGlyThrLeuGluPro 97
DB 232 GCCCCTCTCTGGGTCTCTCCGCTGCTCACTGTTTCATGACGAATGGAGCGCTGGAGCC 291
QY 98 AlaAlaGluTrpSerValLeuLeuGlyValHisSerGlnAspGlyProLeuAspGlyAla 117
DB 292 GCGGCGGAGTGTGCTACTGCTGGCGGTGCACTCCAGAGCGGGCCCTGGAGCGGCG 351
QY 118 HisThrArgAlaValAlaAlaValValProAlaAsnTyrSerGlnValGluLeuGly 137
DB 352 CACACCGCGCAGTGGCGCCATCGTGGTGGCGGCACTACAGCCAGTGGAGCTGGGC 411
QY 138 AlaAspLeuAlaLeuLeuArgLeuAlaSerProAlaSerLeuGlyProAlaValTrpPro 157
DB 412 GCGACCTGGCCCTGCTGGCGCTGGCTCACCAGCAGCTGGGCCCGCGCTGGGCT 471
QY 158 ValCysLeuProArgAlaSerHisArgPheValHisGlyThrAlaCysTrpAlaThrGly 177
DB 472 GTCTGCTGCTGGCGCCCTCACACCGCTTCTGTCAGCGGACCGCTGCTGGGCGCAC 531
QY 178 TrpGlyAspValGlnGluAlaAspProLeuProLeuProTrpValLeuGlnGluValGlu 197
DB 532 TGGGGAGAGCTGCAGAGAGCAGATCTCTGCTGCTCTCCCTGCTGCTGCTGCTGCTGCT 591
QY 198 LeuArgLeuLeuGlyGluAlaThrCysGlnCysLeuTyrSerGlnProGlyProPheAsn 217
DB 592 CTAAAGCTGCTGGCGGAGCCACTCTCAATGCTCTACAGCCAGCCCGCTCCCTTCAAC 651
QY 218 LeuThrLeuGlnLeuLeuProGlyMetLeuCysAlaGlyTyrProGluGlyArgArgAsp 237
DB 652 CTCACTCTCAGATATTGCCAGGAGTCTGTGTGCTGGCTACCCAGAGGGCGCGCAGGAC 711
QY 238 ThrCysGlnCysAspSerGlyGlyProLeuValCysGluGluGlyGlyArgTrpPheGln 257
DB 712 ACCTGCCAGGTGACTCTGGGGGGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 771
QY 258 AlaGlyLeuThrSerPheGlyPheGlyCysGlyArgArgAsnArgProGlyValPheThr 277
DB 772 GCAGGAATCACCAGCTTGGCTTGGCTGTGGAGCGGAGAAACCGCTGGAGTTTCACT 831
QY 278 AlaValAlaThrTyrGluAlaTrpIleArgGluGlnValMetGlySerGluProGlyPro 297
DB 832 GCTGTGGTACCTATGAGGATGATACGGGAGCAGGTGATGGGTTCAGAGCGCTGGGCT 891
QY 298 AlaPheProThrGlnProGlnLysThrGlnSerAsp 309
DB 892 GCCTTTCCACCCAGCCCGCAGAGACCCAGTCAGAT 927

RESULT 2
US-10-221-097-10
; Sequence 10, Application US/10221097
; Publication No. US20030144476A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj

; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoxing
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50016
; CURRENT APPLICATION NUMBER: US/10/221,097
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: PCT/US01/07143
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/187,107
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: 60/236,874
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/188,916
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/237,846
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 768
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-221-097-10

Alignment Scores:
Pred. No.: 3.59e-35 Length: 768
Score: 45.00 Matches: 45
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.61% Indels: 0
DB: 14 Gaps: 0

US-10-037-417-46 (1-357) x US-10-221-097-10 (1-768)

QY 47 IleValGlyGlySerAsnAlaGlnProGlyThrTrpProTrpGlnValSerLeuHisHis 66
DB 1 ATCGTGGGGGCTCAACGCGCAGCGGGCAGCTGGCCCTGGCAAGTGGAGCTGGACCAT 60
QY 67 GlyGlyGlyHisLeCysGlyGlySerLeuLeuAlaProSerTrpValLeuSerAlaAla 86
DB 61 GGAGGTGGCCACATCGCGGGGCTCCCTCATGCCCGCTCTCTGGGTCTCTCCGCTGCT 120
QY 87 HisCysPheMetThr 91
DB 121 CACTGTTTCATGACG 135

RESULT 3
US-10-029-386-21306/c
; Sequence 21306, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21306
; LENGTH: 315
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
OTHER INFORMATION: NT HIT: Q16651, EVALUE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: Q16651, EVALUE 1.00e-08
OTHER INFORMATION: EST_HUMAN HIT: AUI42128.1, EVALUE 0.00e+00
US-10-029-386-21306

Alignment Scores:
Pred. No.: 1.18e-06 Length: 315
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 14 Gaps: 0

US-10-037-417-46 (1-357) x US-10-029-386-21306 (1-315)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
|||||
Db 86 ATGCCCCAGAGGGGCTCTGGGGCTGGCGAGCTGGGGGCTGTGGCC 39

RESULT 4

US-10-029-386-7606/c

Sequence 7606, Application US/10029386

Publication No. US20030194704A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G

FILE REFERENCE: ABOICA-X-2

CURRENT APPLICATION NUMBER: US/10/029,386

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 34288

SOFTWARE: Annonax Sequence Listing Engine vers. 1.1

SEQ ID NO 7606

LENGTH: 543

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO CHR16.3

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.3

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6

OTHER INFORMATION: SWISSPROT HIT: Q16651, EVALUE 7.00e-04

OTHER INFORMATION: EST_HUMAN HIT: AUI42128.1, EVALUE 0.00e+00

OTHER INFORMATION: NT HIT: U33446.1, EVALUE 0.00e+00

US-10-029-386-7606

Alignment Scores:
Pred. No.: 1.87e-06 Length: 543
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 14 Gaps: 0

US-10-037-417-46 (1-357) x US-10-029-386-7606 (1-543)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
|||||
Db 209 ATGCCCCAGAGGGGCTCTGGGGCTGGCGAGCTGGGGGCTGTGGCC 162

RESULT 5

US-10-042-865-31

Sequence 31, Application US/10042865

Publication No. US20040029216A1

GENERAL INFORMATION:

APPLICANT: Padigaru, Muralidhara

APPLICANT: Li, Li

APPLICANT: Zerhusen, Bryan D

APPLICANT: Li, Li
APPLICANT: Zerhusen, Bryan D
APPLICANT: Casman, Stacie J
APPLICANT: Shenoy, Suresh G
APPLICANT: Spytek, Kimberly
APPLICANT: Zhong, Mei
APPLICANT: Gangoli, Esha A
APPLICANT: Burgess, Catherine E
APPLICANT: Patturajan, Meera
APPLICANT: Vernet, Corine A.M
APPLICANT: Taylor, Sarah
APPLICANT: Tchernev, Velizar T
APPLICANT: Miller, Charles E
APPLICANT: Guo, Xiaojia
APPLICANT: Boldog, Ference L
APPLICANT: Grosse, William M
APPLICANT: Alsobrook II, John P
APPLICANT: Gerlach, Valerie L
APPLICANT: Edinger, Shlomit R
APPLICANT: Rothenberg, Mark E
APPLICANT: Ellerman, Karen
APPLICANT: MacDougall, John
APPLICANT: Malyankar, Uriel M
APPLICANT: Millet, Isabelle
APPLICANT: Peyman, John
APPLICANT: Smithson, Glemnda
APPLICANT: Gunther, Erik
APPLICANT: Stone, David
TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
FILE REFERENCE: 21402-537
CURRENT APPLICATION NUMBER: US/10/042,865
CURRENT FILING DATE: 2002-05-17
PRIOR APPLICATION NUMBER: 60/260,417
PRIOR FILING DATE: 2001-01-09
PRIOR APPLICATION NUMBER: 60/260,831
PRIOR FILING DATE: 2001-01-10
PRIOR APPLICATION NUMBER: 60/272,338
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/274,876
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/284,704
PRIOR FILING DATE: 2001-04-18
NUMBER OF SEQ ID NOS: 264
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 31
LENGTH: 1161
TYPE: DNA
ORGANISM: Homo sapiens
US-10-042-865-31

Alignment Scores:
Pred. No.: 3.56e-06 Length: 1161
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 12 Gaps: 0

US-10-037-417-46 (1-357) x US-10-042-865-31 (1-1161)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
|||||
Db 1 ATGCCCCAGAGGGGCTCTGGGGCTGGCGAGCTGGGGGCTGTGGCC 48

```
; APPLICANT: Casman, Stacie J
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zhong, Mei
; APPLICANT: Gangolli, Esha A
; APPLICANT: Burgess, Catherine E
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine A.M
; APPLICANT: Taylor, Sarah
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Miller, Charles E
; APPLICANT: Guo, Xiaojia
; APPLICANT: Boldog, Ference L
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Gerlach, Valerie L
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Rotherberg, Mark E
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John
; APPLICANT: Smithson, Glennda
; APPLICANT: Gunther, Erik
; APPLICANT: Stone, David
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-537
; CURRENT APPLICATION NUMBER: US/10/042,865
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/260,417
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/260,831
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/272,338
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/274,876
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/284,704
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 264
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 1726
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-042-865-29

Alignment Scores:
Pred. No.: 4.98e-06 Length: 1726
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 12 Gaps: 0

US-10-037-417-46 (1-357) x US-10-042-865-29 (1-1726)
Qy 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 229 ATGCCCAAGAGGGGGTCTCTGGGGCTGGGCAGCTGGGGGCTGTGGCC 276

RESULT 7
US-10-176-847-85
; Sequence 85, Application US/10176847
; Publication No. US2003006836A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/09/880,107

US-10-037-417-46 (1-357) x US-10-176-847-85 (1-1733)
Qy 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 99 ATGCCCAAGAGGGGGTCTCTGGGGCTGGGCAGCTGGGGGCTGTGGCC 146

RESULT 8
US-09-948-094-1
; Sequence 1, Application US/09948094
; Patent No. US20020090625A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Mok, Samuel
; APPLICANT: Wong, Kwong-kwok
; TITLE OF INVENTION: Methods of Detecting Cancer Based on Prostatin
; FILE REFERENCE: 81994/282423
; CURRENT APPLICATION NUMBER: US/09/948,094
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (229)..(1260)
; US-09-948-094-1

Alignment Scores:
Pred. No.: 5.25e-06 Length: 1834
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 9 Gaps: 0

US-10-037-417-46 (1-357) x US-09-948-094-1 (1-1834)
Qy 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 229 ATGCCCAAGAGGGGGTCTCTGGGGCTGGGCAGCTGGGGGCTGTGGCC 276

RESULT 9
US-09-880-107-2214
; Sequence 2214, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
```

APPLICANT: John MONAHAN
APPLICANT: Manjula GANNAVAPURU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shubhangi KAMATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Rachel E. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Peter VEIBY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumei ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
FILE REFERENCE: Assessment, Prevention, and Therapy of Ovarian Cancer
CURRENT APPLICATION NUMBER: US/10/097,340
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 261
LENGTH: 1834
TYPE: DNA
ORGANISM: Homo sapiens
US-10-097-340-261

Alignment Scores:
Pred. No.: 5.25e-06 Length: 1834
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 9 Gaps: 0
US-10-037-417-46 (1-357) x US-09-880-107-2214 (1-1834)

1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276

RESULT 10
US-09-967-768A-141
Sequence 141, Application US/09967768A
Patent No. US20020150677A1
GENERAL INFORMATION:
APPLICANT: Augustus, Meena
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
FILE REFERENCE: Sets
CURRENT APPLICATION NUMBER: US/09/967,768A
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: US/60/236,109
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,034
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,111
PRIOR FILING DATE: 2000-09-28
NUMBER OF SEQ ID NOS: 325
SOFTWARE: PatentIn version 3.0
SEQ ID NO 141
LENGTH: 1834
TYPE: DNA
ORGANISM: Homo sapiens
US-09-967-768A-141

Alignment Scores:
Pred. No.: 5.25e-06 Length: 1834
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 9 Gaps: 0
US-10-037-417-46 (1-357) x US-09-967-768A-141 (1-1834)

1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276

US-09-967-768A-141
Sequence 141, Application US/09967768A
Patent No. US20020150677A1
GENERAL INFORMATION:
APPLICANT: Augustus, Meena
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
FILE REFERENCE: Sets
CURRENT APPLICATION NUMBER: US/09/967,768A
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: US/60/236,109
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,034
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,111
PRIOR FILING DATE: 2000-09-28
NUMBER OF SEQ ID NOS: 325
SOFTWARE: PatentIn version 3.0
SEQ ID NO 141
LENGTH: 1834
TYPE: DNA
ORGANISM: Homo sapiens
US-09-967-768A-141

Alignment Scores:
Pred. No.: 5.25e-06 Length: 1834
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 9 Gaps: 0
US-10-037-417-46 (1-357) x US-09-967-768A-141 (1-1834)

1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276

US-10-037-417-46 (1-357) x US-09-967-768A-141 (1-1834)
QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276
RESULT 11
US-10-097-340-261
Sequence 261, Application US/10097340
Publication No. US20030087250A1
GENERAL INFORMATION:

Alignment Scores:
Pred. No.: 5.25e-06 Length: 1834
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 9 Gaps: 0
US-10-037-417-46 (1-357) x US-10-097-340-261 (1-1834)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276
RESULT 12
US-10-101-510-447
Sequence 447, Application US/10101510
Publication No. US20030148295A1
GENERAL INFORMATION:
APPLICANT: WAN, JACKSON
APPLICANT: WANG, YIXIN
TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
FILE REFERENCE: 15117.0012
CURRENT APPLICATION NUMBER: US/10/101,510
CURRENT FILING DATE: 2002-03-20
PRIOR APPLICATION NUMBER: 60/276,947
PRIOR FILING DATE: 2001-03-20
NUMBER OF SEQ ID NOS: 805
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 447

Alignment Scores:
Pred. No.: 5.25e-06 Length: 1834
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48% Indels: 0
DB: 14 Gaps: 0
US-10-037-417-46 (1-357) x US-10-097-340-261 (1-1834)

QY 1 MetAlaGlnLysGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 229 ATGGCCCAAGAGGGGTCTTGGGGCTTGGCAGCTGGGGGCTGTGGCC 276
RESULT 12
US-10-101-510-447
Sequence 447, Application US/10101510
Publication No. US20030148295A1
GENERAL INFORMATION:
APPLICANT: WAN, JACKSON
APPLICANT: WANG, YIXIN
TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
FILE REFERENCE: 15117.0012
CURRENT APPLICATION NUMBER: US/10/101,510
CURRENT FILING DATE: 2002-03-20
PRIOR APPLICATION NUMBER: 60/276,947
PRIOR FILING DATE: 2001-03-20
NUMBER OF SEQ ID NOS: 805
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 447


```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 621
; LENGTH: 988
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-621

Alignment Scores:
Pred. No.: 0.00299      Length: 988
Score: 13.00           Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64%      Indels: 0
DB: 14                 Gaps: 0

US-10-037-417-46 (1-357) x US-10-101-510-621 (1-988)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeuValCys 249
Db 770 GACACGTGCCAGGGCGATTCTGGAGGCCCCCTGGTGTGT 808

RESULT 15
US-10-101-510-612
; Sequence 612, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 612
; LENGTH: 994
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-612

Alignment Scores:
Pred. No.: 0.00301      Length: 994
Score: 13.00           Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64%      Indels: 0
DB: 14                 Gaps: 0

US-10-037-417-46 (1-357) x US-10-101-510-612 (1-994)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeuValCys 249
Db 770 GACACGTGCCAGGGCGATTCTGGAGGCCCCCTGGTGTGT 808

Search completed: February 26, 2004, 02:47:28
Job time : 406 secs
```

```
; LENGTH: 3382
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-447

Alignment Scores:
Pred. No.: 8.81e-06      Length: 3382
Score: 16.00           Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.48%      Indels: 0
DB: 14                 Gaps: 0

US-10-037-417-46 (1-357) x US-10-101-510-447 (1-3382)
QY 1 MetaAlaGlnYsGlyValLeuGlyProGlyGlnLeuGlyAlaValAla 16
Db 230 ATGCCCCAGAGGGGTCTCTGGGGCCTGGGCGACGTGGGGCTGTGGCC 277

RESULT 13
US-10-325-745-3
; Sequence 3, Application US/10325745
; Publication No. US20030135028A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Human Serine Protease and Serpin Polypeptides
; FILE REFERENCE: PF391D1
; CURRENT APPLICATION NUMBER: US/10/325,745
; CURRENT FILING DATE: 2002-12-23
; PRIOR APPLICATION NUMBER: US 09/244,111
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: US 60/073,961
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (85)..(708)
; OTHER INFORMATION:
US-10-325-745-3

Alignment Scores:
Pred. No.: 0.00248      Length: 792
Score: 13.00           Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64%      Indels: 0
DB: 14                 Gaps: 0

US-10-037-417-46 (1-357) x US-10-325-745-3 (1-792)
QY 237 AspThrCysGlnGlyAspSerGlyGlyProLeuValCys 249
Db 541 GACACGTGCCAGGGCGATTCTGGAGGCCCCCTGGTGTGT 579

RESULT 14
US-10-101-510-621
; Sequence 621, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
```

AL193077	LOCUS	360 bp	mRNA	linear	EST 30-MAR-1999
AL193077	DEFINITION	tg25f08.x1 NCI CGAP CLL1 Homo sapiens cDNA clone IMAGE:2109831 3', similar to SW:PS8_HUMAN Q16651 PROSTASIN PRECURSOR ;, mRNA sequence.			
ACCESSION	AL193077	GI:4222624			
VERSION	AL193077.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 360)				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapsb@mail.nih.gov Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html Insert Length: 691 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 334. Location/Qualifiers				
FEATURES	source	1..360			
		/organism="Homo sapiens"			
		/mol_type="mRNA"			
		/db_xref="taxon:9606"			
		/clone="IMAGE:2109831"			
		/tissue_type="B-cell, chronic lymphocytic leukemia"			
		/lab_host="DH10B"			
		/clone_lib="NCI_CGAP CLL1"			
		/note="vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGCGCGCGCATTCGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."			
ORIGIN		6.1%; Score 67; DB 9; Length 360;			
	Query Match	100.0%; Pred. No. 1.9e-20;			
	Best Local Similarity	Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
	Matches 57;				
QY	1	GGGCCCTTGCTCTGGCCATGCCCAAGAGGGGCTCTCTGGGGCTGGCGACGTGGGGGCT 60			
Db	127	GGGCCCTTGCTCTGGGGCCATGCCCAAGAGGGGCTCTCTGGGGCTGGCGACGTGGGGGCT 186			
QY	61	GTGGCCA 67			
Db	187	GTGGCCA 193			
RESULT 7					
AL1623099	LOCUS	366 bp	mRNA	linear	EST 15-DEC-1999
AL1623099	DEFINITION	tu49f05.x1 NCI CGAP Pr-28 Homo sapiens cDNA clone IMAGE:2254401 3', similar to SW:PS8_HUMAN Q16651 PROSTASIN PRECURSOR ;, mRNA sequence.			
ACCESSION	AL1623099				
VERSION	AL1623099.1	GI:4648024			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				

BDB 218 GTGGCCA 224
 RESULT 5
 BM828821 318 bp mRNA linear EST 06-MAR-2002
 LOCUS X-EST0101694 S9SNU601 Homo sapiens cDNA clone S9SNU601-51-E07 5',
 DEFINITION mRNA sequence.
 BM828821
 BM828821.1 GI:19185230
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 318)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 51 row: E column: 07
 High quality sequence stop: 318.
 Location/Qualifiers
 1. 318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="SNU-601"
 /lab_host="Top10F"
 /clone_lib="S9SNU601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
 Site 2: XhoI; The poly (A) + RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including SfiI
 site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized with Superscript II using SfiI
 oligo-dT primer. After first strand synthesis RNA was
 degraded by NaOH treatment and cDNA was amplified by PCR
 reaction. The PCR products were digested with SfiI and
 cloned into DrallI- digested pME18S-FL3 vector. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

1. 318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="SNU-601"
 /lab_host="Top10F"
 /clone_lib="S9SNU601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
 Site 2: XhoI; The poly (A) + RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including SfiI
 site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized with Superscript II using SfiI
 oligo-dT primer. After first strand synthesis RNA was
 degraded by NaOH treatment and cDNA was amplified by PCR
 reaction. The PCR products were digested with SfiI and
 cloned into DrallI- digested pME18S-FL3 vector. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

1. 318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="SNU-601"
 /lab_host="Top10F"
 /clone_lib="S9SNU601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
 Site 2: XhoI; The poly (A) + RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including SfiI
 site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized with Superscript II using SfiI
 oligo-dT primer. After first strand synthesis RNA was
 degraded by NaOH treatment and cDNA was amplified by PCR
 reaction. The PCR products were digested with SfiI and
 cloned into DrallI- digested pME18S-FL3 vector. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

1. 318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="SNU-601"
 /lab_host="Top10F"
 /clone_lib="S9SNU601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
 Site 2: XhoI; The poly (A) + RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including SfiI
 site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized with Superscript II using SfiI
 oligo-dT primer. After first strand synthesis RNA was
 degraded by NaOH treatment and cDNA was amplified by PCR
 reaction. The PCR products were digested with SfiI and
 cloned into DrallI- digested pME18S-FL3 vector. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

1. 318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="SNU-601"
 /lab_host="Top10F"
 /clone_lib="S9SNU601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
 Site 2: XhoI; The poly (A) + RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including SfiI
 site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized with Superscript II using SfiI
 oligo-dT primer. After first strand synthesis RNA was
 degraded by NaOH treatment and cDNA was amplified by PCR
 reaction. The PCR products were digested with SfiI and
 cloned into DrallI- digested pME18S-FL3 vector. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

1. 318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="SNU-601"
 /lab_host="Top10F"
 /clone_lib="S9SNU601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
 Site 2: XhoI; The poly (A) + RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including SfiI
 site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized with Superscript II using SfiI
 oligo-dT primer. After first strand synthesis RNA was
 degraded by NaOH treatment and cDNA was amplified by PCR
 reaction. The PCR products were digested with SfiI and
 cloned into DrallI- digested pME18S-FL3 vector. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

1. 318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="SNU-601"
 /lab_host="Top10F"
 /clone_lib="S9SNU601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
 Site 2: XhoI; The poly (A) + RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including SfiI
 site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized with Superscript II using SfiI
 oligo-dT primer. After first strand synthesis RNA was
 degraded by NaOH treatment and cDNA was amplified by PCR
 reaction. The PCR products were digested with SfiI and
 cloned into DrallI- digested pME18S-FL3 vector. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

1. 318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="SNU-601"
 /lab_host="Top10F"
 /clone_lib="S9SNU601"
 /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
 Site 2: XhoI; The poly (A) + RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including SfiI
 site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized with Superscript II using SfiI
 oligo-dT primer. After first strand synthesis RNA was
 degraded by NaOH treatment and cDNA was amplified by PCR
 reaction. The PCR products were digested with SfiI and
 cloned into DrallI- digested pME18S-FL3 vector. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

1. 318
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S9SNU601-51-E07"
 /sex="M"
 /tissue_type="Ascites"
 /cell_type="Epithelial"
 /cell_line="SNU-601"
 /lab_host="Top10

